A Novelty Detection Approach to Seizure Analysis from Intracranial EEG

A Dissertation
Presented to
The Academic Faculty

By

Andrew B. Gardner

In Partial Fulfillment

Of the Requirements for the Degree

Doctor of Philosophy in Electrical Engineering

School of Electrical and Computer Engineering

Georgia Institute of Technology

Atlanta, GA USA

April 2004

A Novelty Detection Approach to Seizure Analysis from Intracranial EEG

Approved by:

Dr. Aaron Lanterman	Dr. Robert Butera
Dr. Rosana Esteller	Dr. Brian Litt
Dr. Arthur Koblasz	Dr. George Vachtsevanos

April 8, 2004

Dedication

To my family and friends—especially to my father, for taking the long view.

Acknowledgements

I owe special thanks to Drs. Javier Echauz, Rosana Esteller, Arthur Koblasz, Brian Litt, and George Vachtsevanos. Their unflagging optimism, guidance, technical expertise, and general support were the bedrock of this work. I extend special thanks to my advisors, Drs. Brian Litt and George Vachtsevanos, for offering valuable insights at all of the key moments in the development of this work, sharing their support and passion for this research, and providing an environment in which I might succeed. I thank Maryann D'Alessandro for helping me in the early stages of this research, and shaping my research interests. I thank all of the other members of the group, past and present, for their wonderful discussions and advice along the way. I also thank the members of my committee, Drs. Robert Butera, Rosana Esteller, and Aaron Lanterman, for their time, effort, and early suggestions necessary to complete this work.

I thank my father and mother, Walter and Sylvia Gardner, for encouraging my efforts, even when it wasn't clear to me where they would lead. I thank my siblings, Melissa, Kara, and Eli for their belief in me. I owe special thanks to two friends: Bob Copenhaver and Edgar Jones. Their support, encouragement, and mentoring maintained me in this work, and set me on the path to success.

Finally, I would like to thank Martha Duffy. She has been ever supportive, during the best and worst of it.

Table of Contents

List of Figures

List of Tables

Summary

The objective of this research is to create a general framework for support vector machine classification of time series events, and to apply that framework to analyze physiological signals recorded from epileptic patients. In contrast to previous works, this research formulates seizure analysis as a novelty detection problem. This approach allows seizure detection and prediction to be treated uniformly, in a way that is capable of accommodating multichannel and/or multimodal measurements. Theoretical properties of the support vector machine algorithm employed provide a straightforward means for controlling the false alarm rate of the detector. The resulting novelty detection system was evaluated both offline and online on a corpus of 1077 hours of intracranial electroencephalogram (IEEG) recordings from 12 patients diagnosed with medically resistant temporal lobe epilepsy during evaluation for epilepsy surgery. These patients collectively had 118 seizures during the recording period. The performance of the novelty detection framework was assessed with an emphasis on four key metrics of seizure detection: (1) sensitivity (probability of correct detection), (2) mean detection latency, (3) early-detection fraction (prediction or detection of seizure prior to electrographic onset), and (4) false positive rate. Both the offline and online novelty detectors achieved state-of-the-art seizure detection performance. In particular, the online detector achieved 97.85% sensitivity with an average of 1.74 false positive predictions per hour (Fph). Over 40% of the seizures detected were identified prior to their electrographic (EEG) onset, as determined by an epileptologist. The mean and

median detection latencies for the online detector were -13.3 seconds and 6.00 seconds. These results demonstrate that a novelty detection approach is not only feasible for seizure analysis, but it improves upon the state-of-the-art as an effective, robust technique for detecting these events and provides a method for predicting 40% of seizures. Additionally, an extension of the basic novelty detection framework demonstrated its use as a simple, effective tool for examining the spread of seizure onsets. This may be useful for automatically identifying seizure focus channels in patients with focal epilepsies. It is anticipated that this research will aid in localizing seizure onsets, and provide more efficient algorithms for use in a real device.

Chapter 1

Introduction

1.1 Motivation

Epilepsy, a neurological disorder in which patients suffer from recurring seizures, affects approximately 1% of the world population. In the United States, 200,000 new cases are reported annually. There are more than 30 distinct classes of seizure. Their manifestations range from subtle, abnormal sensations to unpredictable changes in conscious state, to immediate loss of consciousness and convulsions. In spite of available dietary, drug, and surgical treatment options, more than 25% of individuals with epilepsy have seizures that are uncontrollable [1]. Daily life for these patients is greatly impaired—education, employment, and even transportation can become difficult endeavors. Many new therapies for medically resistant epilepsy are being investigated. Among the most promising are implantable devices that deliver local therapy, such as direct electrical stimulation or chemical infusions, to affected regions of the brain. These treatments will rely on robust algorithms for seizure detection and prediction to perform effectively.

While the specific cause of epilepsy in an individual is often unknown and the mechanisms that cause seizure onset are not well understood, several successful

techniques for detecting—and even modestly predicting—seizures have been published. All reported approaches, however, suffer from one or more of the following limitations:

- Accurate detection/prediction requires careful, patient-specific tuning
- Seizure detections do not occur "early-enough" (long detection delay)
- Seizure prediction accuracy is highly variable and often marginally better than 50%
- *A priori* focus channel identification is required
- Usefulness for poorly localized epilepsies is limited
- Seizure data (which is expensive to collect) is required for training

Techniques for overcoming these limitations hold promise for more precise and widely applicable methods to control or eliminate seizures. This research develops one technique for improving the state-of-the-art in seizure detection and prediction that addresses all of these issues. The result of this work is applicable to many other biomedical and engineering problems involving event detection and prediction.

1.2 Terminology

A *time series*, or *signal*, $y = \{(t_k, x_k)\}_k$, is an ordered collection of tuples with each observation, x_k, made at a corresponding time instant, t_k. When t_k is discrete, we refer to the signal as *discrete time*. When the time interval between successive observations is constant the signal is *uniformly sampled*. A *time-frequency distribution (TFD)* simultaneously characterizes a signal's time and frequency components.

Machine learning refers collectively to algorithms that produce regression, classification, density modeling, or clustering models. In regression (i.e., "curve-fitting") a fitting function, $f : X \subseteq \mathbb{R}^n \to Y \subseteq \mathbb{R}$, is determined. The goodness-of-fit is measured by an objective function. Common objective functions include mean square error and median square error. Classification (i.e., "labeling") refers to the special case of regression where $Y \subseteq \{1,2,\ldots\}$. A *supervised learning* technique is one where both the inputs and desired outputs are known. Regression and classification are supervised learning techniques.

Seizure analysis refers collectively to algorithms for seizure detection, seizure prediction, and automatic focus channel identification. These analyses are primarily performed on the *electroencephalogram (EEG)*. The EEG is a measure of brain electrical activity. The following EEG types are differentiated by electrode location: the *electrocorticogram (ECoG)*, recorded directly from the cortex, usually from subdural electrode strips or grids; *depth EEG*, recorded from inside the brain using penetrating electrodes; the *intracranial EEG (IEEG)*, recorded from anywhere inside the skull; and the *scalp EEG*, recorded from the surface of the head.

For seizure analysis it is convenient to describe segments of physiological signals by their temporal proximity to seizure activity. A *segment* or *epoch* refers to a continuous interval of EEG recording. The *ictal* period refers to the time during which a seizure occurs. The *interictal* period is the time between successive seizures. *Preictal*- and *postical*-periods refer to the times immediately preceding and following a seizure, respectively. *Baseline* refers to recording intervals that are three or more hours away from a seizure onset. Additional terminology refers to specific seizure-related events: the

unequivocal electrographic onset (UEO) is defined as the earliest time that a seizure occurrence is evident to an epileptologist viewing an EEG without prior knowledge that a seizure follows; the *earliest electrographic change (EEC)* is defined as the earliest time that abnormal, seizure-related EEG activity is evident to an epileptologist viewing an EEG given the prior knowledge that a seizure soon follows; the *unequivocal clinical onset (UCO)* is the earliest time that a seizure occurrence is apparent by visually observing a patient. *Seizure onset* in this work is synonymous with UEO. Figure 1-1 shows a seizure and related terminology for a two-channel EEG recording.

General epilepsy-related terminology includes: *ictogenesis*, the process by which a seizure is generated; *prodrome*, a clinical sensation that a seizure may occur soon; *aura*, the sensation associated with a very focal seizure; *precursor*, a signature event, typically electrographic, associated with an impending seizure; *focus channel*, the electrode location that exhibits the earliest evidence of seizure activity or, if this is simultaneous in more than one channel, the channel in which the activity is maximal in amplitude.

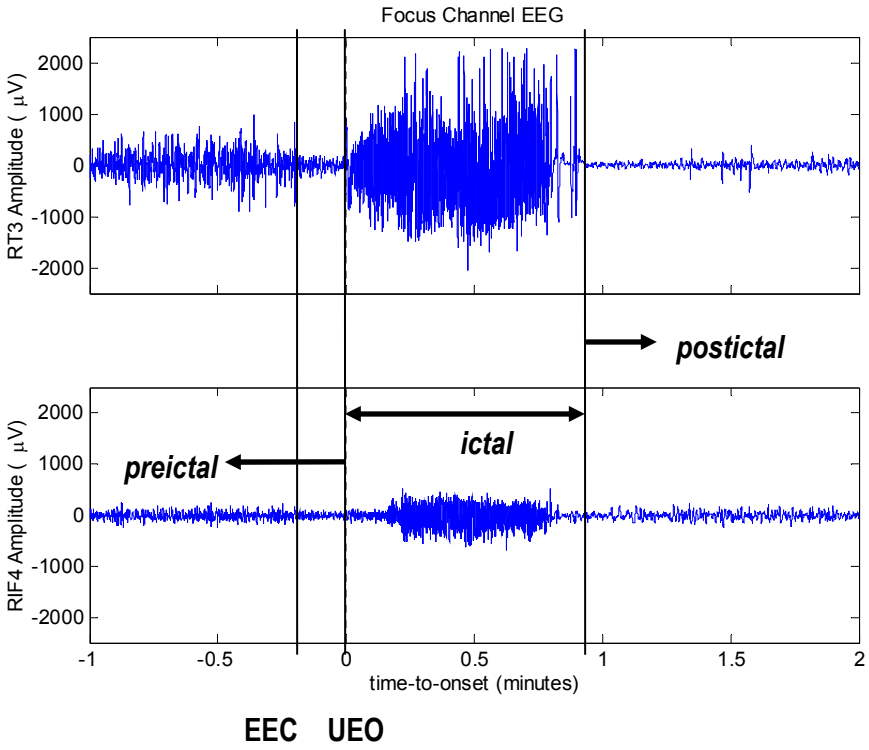

Fig. 1-1. (*Top*) Focus channel EEG showing a seizure; (*Bottom*) simultaneous recording from a non-focus channel. Seizure-related terminology is illustrated.

To characterize detector performance, temporal relationships between ictal state and the detector output must be considered. The four relationships of interest in this work are:

1. *True Positive (TP),* also referred to as a *block true positive* – the detector declares a novelty interval that overlaps with a seizure interval.

2. *False Positive (FP),* also referred to as a *block false positive* – the detector declares a novelty interval, but the IEEG does not exhibit a seizure.
3. *False Negative (FN),* also referred to as a *block false negative* – the detector declares an interval as normal, but that interval overlaps with a seizure interval.
4. *True Negative (TN),* also referred to as a *block true negative* – the detector declares an interval as normal, and that interval exhibits no ictal activity.

Figure 1-2 illustrates these relationships. Additionally, the *detection latency*, τ, is defined to characterize the time lag associated with a seizure detection. It is computed as the time difference between the leading edge of the novelty interval and the seizure onset. Negative values for detection latency indicate early-detections (or predictions) of seizure onsets.

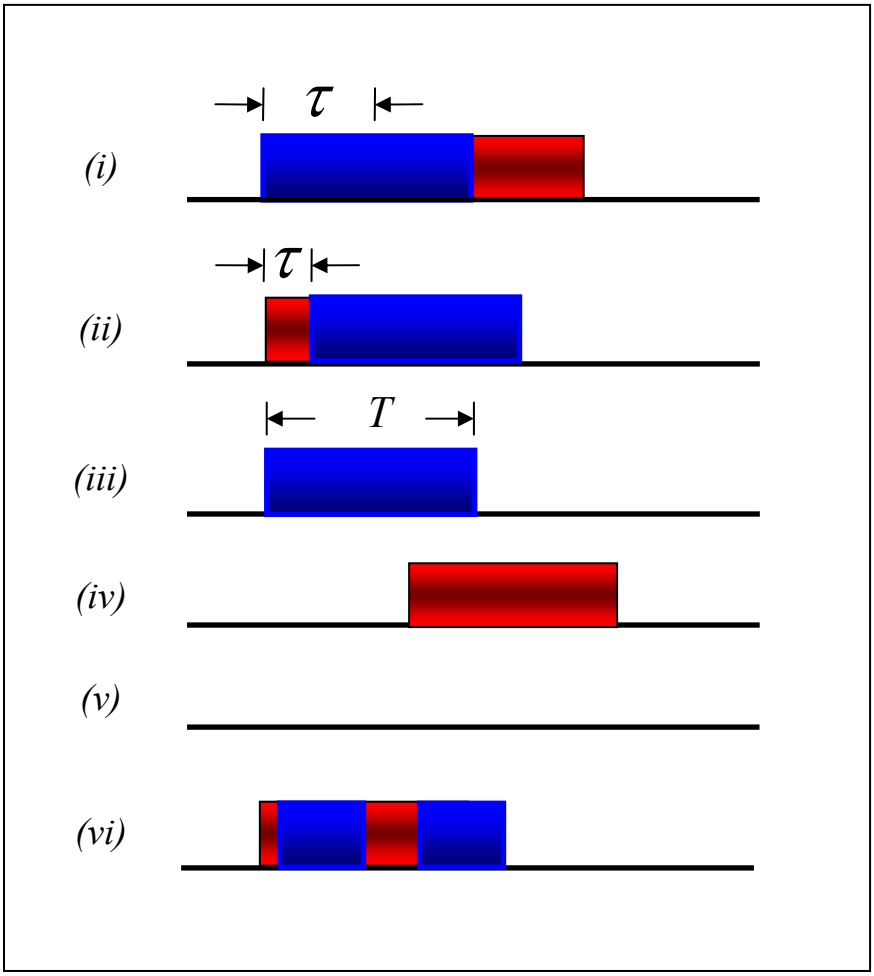

Fig. 1-2. Temporal relationships considered in detector evaluation. Blue intervals represent detected novelty. Red intervals represent ictal activity. *(i), (ii)* Two examples of true positives (e.g., the novelty output interval overlaps ictal activity). The detection latency, τ, is also shown. An early-detection results in negative latency. *(iii)* A false positive detector error. *(iv)* A false negative detector error. *(v)* A true negative. *(vi)* An example of a degenerate case (multiple detection) producing both a true positive and a false positive event.

1.3 Problem Statement

The objective of the proposed research was to create a general framework for *support vector machine (SVM)* classification of time series events, and to apply that

framework to the analysis of physiological signals recorded from epileptic patients for seizure analysis. In contrast to previous work, this research considered an alternative formulation of seizure analysis as a novelty detection problem. This approach offers a unified treatment of seizure detection and prediction while making few assumptions regarding ictogenesis. Theoretical properties of the SVM algorithm employed provide a simple means for controlling the false alarm rate. Straightforward extensions to the basic approach accommodate multichannel and multimodal features. Of note is that this method requires no seizure data for training. Within this context, the following research aims were addressed:

1. A novelty detection framework was developed and tested to demonstrate state-of-the-art performance at seizure analysis.
2. The effects of model parameter variations on the performance of the novelty detection framework were investigated.
3. A naïve online novelty detector framework was implemented and evaluated on a corpus of more than 1077 hours of IEEG data to demonstrate state-of-the-art seizure analysis performance.
4. An extension to the framework was developed to allow investigation of seizure onsets, and possible automatic determination of focus channel.

1.4 Background and Literature Review

Previous works summarize seizure detection [2, 3] and prediction [4, 5] in detail. A brief treatment is repeated here from the perspective of this research. A focused summary of relevant work on SVMs concludes this section.

1.4.1 Seizure Detection

Early attempts to detect seizures began in the 1970s [6, 7] and primarily considered scalp EEG recordings to detect the clinical (and less frequently) electrographic onset of seizures. Beginning in the 1970s, Gotman began publishing papers on the automatic analysis of EEG waveforms. In 1983, he looked at coherence to measure interchannel differences in onset times to study seizure propagation [3]. A consequence of this work was an analysis technique for automated seizure detection. In 1990, he reported results for 293 seizure recordings from 49 patients [8]. The reported average false alarm rate was one *false positive per hour (Fph)*; 24% of all seizures went undetected. The false detection problem was then attacked by training a classifier to recognize false alarms. In 1993, Gotman reported that the false detection rate for this method fell 33% [9]. By 1995, an early seizure warning system, trained on template EEG activity, was presented along with impressive detection results: 100% accuracy with an average latency of 9.35 seconds and a false alarm rate of 0.02 Fph [10]. A later publication [11] reported similar results (9.35 second latency) using time- and frequency-domain features classified by a k-nearest neighbor classifier. A total of 47 seizures from 12 patients were considered. Most recently, Zijlmans and Gotman have found that changes in parasympathetic activity (e.g., heart rate) frequently manifest prior to the

earliest EEG evidence of seizure onset [12]. This is consistent with much earlier work qualitatively examining seizure-related heart rate changes (e.g., [13]).

Time-frequency methods for seizure detection have a long research history. In 1993, Zaveri et al. showed that the mean Teager energy of ECoG recordings yielded 100% detection accuracy on an 11-seizure database [14]. Schiff and co-investigators considered a multiresolution analysis of the EEG in 1994 [15] and pulse amplitude models in 1997 [16]. Both investigations examined ECoG recordings and found qualitative seizure-related changes; though no specific detection algorithm was posed. Williams and Zaveri continued their time-frequency analysis in 1995 and introduced the reduced interference distribution (RID) for seizure analysis [17]. They noted the presence of preseizure chirps, but did not present specific detection results. Similar publications the same year reported chirp identification using scalograms and spectrograms [18], and a matched filter approach [19]. In 2001, Zaveri et al. examined 30-second epochs of data with RIDs and suggested instantaneous frequency as a detection feature [20]. Other approaches using wavelets (e.g. [21]) have been reported with similar qualitative results. To date, most time-frequency analyses have concurred on the presence of interesting, preictal time-frequency structures, but no practical detection or prediction algorithms have emerged.

The 1990s saw a marked increase in seizure-related research. A variety of techniques for detection were considered. In 1996, Gabor used self-organizing map neural networks on scalp EEGs to classify 62 seizures without quantitative results [22, 23]. In 1997, Marchesi et al. explored genetic programming for learning to detect spike-and-slow-wave activity in EEG recordings using a variety of simple features [24]. No

validation was performed on seizure data. In 1998, Osorio et al. achieved perfect detection using a measure called seizure intensity. The average detection latency was 2.1 seconds evaluated on a database of 125 patients, but the same data was used for training and validation [25]. In 2001, they qualitatively described a system for real-time detection and intervention with direct electrical stimulation of the brain [26]. More recent seizure intensity results using offline ECoG recordings were reported in 2002 [27].

In 1999, Esteller et al. presented state-of-the-art results on fractal dimension-based seizure detection reporting a mean detection latency of 9.8 seconds [28, 29]. Her thesis evaluated more than 20 features for detection and found that fractal dimension, wavelet packet energy, and Teager energy were the most promising [2]. It marked one of the early attempts to incorporate multiple features for detection and introduced specific terminology for seizure analysis. Results on the use of line length for detection were reported the following year: the average detection latency was reduced to 4.1 seconds with a false alarm rate of 0.051 Fph [30]. A total of 111 seizures (many subclinical) were used for validation. A current detector used by NeuroPace, Inc., claims a detector sensitivity of 97%, a mean detection latency of 5.01 seconds, and a false negative rate of 1.54% at a 0.013 Fph [31]. Their detector was evaluated on 1265 hours of IEEG data, but was tuned heuristically in a patient-specific manner and evaluated offline. Their results represent the current state-of-the-art in seizure detection systems.

Other approaches for detection have included: autoregressive (AR) modeling, rules-based analysis of amplitude and frequency changes of scalp EEG [32], singular-spectrum techniques for scalp EEG [33], and nonlinear techniques like multidimensional probability evolution [34]. Seizure detection also appears in the patent literature. Since

1973, the year the first patent for seizure detection was issued [6], more than 20 other patents on early detection and intervention have been issued.

Over the past 30 years seizure detection technology has matured. Equally important, so have the techniques for validating results. Prior to the mid-1990s, precise terminology for seizure-related events was not in place. Metrics like detection latency were poorly defined, and it was common to read about single-feature oriented detectors. Current research now focuses on improved accuracy, decreased latency, application to poorly localized seizures, and extension to multichannel recordings.

1.4.2 Seizure Prediction

To date, no comprehensive, prospective study of seizure prediction has been conducted. In contrast to seizure detection research, precise terminology and methods for evaluating seizure prediction claims are not well established—even the definition of a seizure is not agreed upon. Much spirited debate surrounds the theories of ictogenesis, the possibility of seizure prediction, and the relationship between seizure detection and prediction. It should be noted that the epilepsy research community rarely shares data, making independent verification of seizure prediction results almost impossible. Still, evidence that seizures are predictable does exist: SPECT imaging [35], studies of parasympathetic nervous activity [36-39], diary entries of epileptic patients (e.g., noting auras) [40], and even seizure-alert dogs [41-44] indicate that seizure-related physiological changes occur minutes or hours before electrographic onset. Litt and Echauz [5] comprehensively review this and other evidence in the context of historical efforts at seizure prediction. This section owes much of its content and organization to their paper.

Viglione reported the first serious attempt at seizure prediction in 1970 [7]. A primitive filter bank energy detector and perceptron classified scalp EEG as preseizure or baseline (preseizure segments were defined as occurring 10 minutes prior to onset). The system predicted with "reasonable accuracy," but suffered from many false positives.

Seizure prediction research in the 1970s and 1980s focused on morphological features (e.g., spikes or bursts) evident in EEGs. Results reported by Gottman [45], Sherwin [46], and Lange [47] showed evidence of increased interictal spiking prior to onset in two studies, but not the third [45]. In two publications ([48, 49]), an examination of the predictability of interictal interval times yielded inconclusive results. More recently, Litt et al. [50] studied long-term energy bursts, subclinical seizure-like bursts (chirps), and accumulated energy features in five patients. They claimed a mean seizure prediction time of 18.5 minutes with three false negatives and six false positives. They concluded that temporal lobe epilepsy was generated in a cascade of electrophysiological events that occurred over minutes to hours. They also postulated that prediction, in the form of identifying periods when the probability of seizure onset is greatly increased, may be possible; however, their results were not comprehensive.

Nonlinear dynamic approaches to seizure prediction have been reported since 1988 [51-55]. Iasemidis, Sackellares, Williams, and Zaveri reported changes in the principal Lyapunov exponent (PLE) for temporal lobe epilepsy. The authors claimed that thresholding the PLE on epochs ranging from seconds to hours predicted seizure onsets 1 – 60 minutes in advance [54], with large interpatient variability of the results. Elger and Lehnertz began reporting results in 1995 on the use of correlation dimension [56]. One

study claimed distinguishable patterns in 15 out of 16 patients that could predict seizures from 4.25 – 25 minutes prior to onset.

In 1998, Geva used fuzzy classification of wavelet features from rat EEGs in a model of generalized convulsive epilepsy, and found predictive indicators four minutes prior to onset [57]. Publications by Petrosian claim the use of EEG texture features for prediction [58-60]. In one instance, a 1.5-minute prediction horizon was claimed from analysis of a single seizure. Poorly defined features and methodology hampered his work. D'Alessandro used a genetic algorithm in conjunction with energy-based features in a patient-tuned architecture [4]. She claimed 62.5% prediction accuracy at 0.28 Fph on 30 seizure segments from patients with temporal lobe epilepsy. The mean prediction horizon was 7.5 minutes. Most interestingly, she found that non-focus channels gave the earliest warning of impending seizure.

As in the early seizure detection work, seizure prediction results are still inconclusive and techniques are in their infancy. Common terminology and practices need to be standardized and data needs to be shared to allow for independent validation of claims.

1.4.3 SVM Classification

Traditional classification architectures rely on empirical risk minimization algorithms to specify "good" models for a classification decision function; as such, they are prone to over- or underfitting. In addition, their performance tends to be highly sensitive to parameter tuning and researcher skill. Statistical learning theory poses a *structural risk minimization* (SRM) criterion that balances the trade-off between good empirical performance (i.e., classification accuracy on training data) and good

generalization ability (i.e., classification accuracy on unseen data). One popular application of SRM is the SVM, first presented in 1992 [61].

The basic idea behind the SVM is to find a hyperplane in a feature space that "optimally" separates two classes. Many other linear learning machines have been considered for this task (e.g., the Perceptron, Fisher's Discriminant, least squares, ridge regression, etc. [62]), but the SVM yields a unique solution that can be shown to minimize the expected risk of misclassifying unseen examples [63]. Training algorithms involve the solution of a well-known optimization problem—constrained quadratic programming—that is computationally efficient and yields global solutions (cf. neural networks which yield local solutions). Several excellent tutorials provide historical context and details on the SVM [64-67].

Extensions to the basic SVM provide rich, nonlinear decision surfaces by first mapping the inputs into a higher-dimensional feature space and performing a linear separation in that feature space. This seems at first counterintuitive because of the "curse of dimensionality:" (1) transforming the input data to a higher-dimensional space can be computationally expensive, and (2) a sparse representation in a high-dimensional feature space increases the potential for under- and overfitting. The first problem is addressed by the well-known kernel trick; the second by the use of the SRM implicit in the SVM formulation [62].

The υ-SVM improves the basic SVM by changing the definition (and hence interpretation) of the penalty parameter for misclassification. Whereas the classic SVM algorithm requires the specification of a nonintuitive misclassification penalty, $C \in [0, \infty)$, the interpretation of $\upsilon \in [0,1]$ is straightforward: (1) it specifies the expected

fraction of misclassified instances, and (2) it specifies the number of support vectors returned by the algorithm. This allows for precise and intuitive control over classification accuracy and computational complexity. Despite this improvement, fewer papers using the υ-SVM have been published, perhaps because of its more rigorous formulation or association with novelty detection applications.

SVMs possess many desirable properties that make them popular in machine learning applications—low-to-moderate computational complexity, a globally optimal training technique, strong theoretical generalization properties, and easy interpretability. These are evident in two of the more commonly cited SVM applications: real-time face detection using simple histograms [68] and text categorization using sparse, high-dimensional feature vectors [69]. Examples of other successful applications span a broad spectrum of engineering problems, including: jet engine fault detection [70], speaker identification [71], network intrusion detection [72], spam filtering [73], and time series prediction [74]. The remainder of this section reviews applications of SVMs to time series analysis, biomedical classification, and novelty detection.

1.4.3.1 SVM Methods for Time Series Analysis

In 1997, SVM regression was evaluated for nonlinear time series prediction [75]. Performance and sensitivity to parameter specification were directly compared to state-of-the-art approaches such as neural networks. Superior performance and minimal parameter sensitivity for SVMs using polynomial and radial basis function kernels were reported. In 1999, Müller et al. studied the effect of different loss functions for SVM time series prediction [74], and Rosipal and Girolami reported that an adaptive SVM regression filter outperformed resource allocating networks [76]. In 2000, Sukens and

Vandewalle presented the recurrent LS-SVM in the context of predicting chaotic time series [77]. A comparison between the LS-SVM and recurrent neural networks concluded that the LS-SVM technique was superior.

Applications of SVMs to object tracking in video have also been presented (e.g., [78, 79]). These approaches combine low-level machine vision techniques with SVM classification to identify blobs. Blob locations are tracked using methods such as Kalman filtering or splining, and updated blob position vectors are used to retrain the classifier. This process is iterated for each frame of video to update the support vectors. Performance is comparable or superior to other object-tracking methods, but the approach is computationally expensive.

Another Kalman-filtering approach for support vector updating was introduced in 1999 [80]. The so-called sequential support vector machine used radial basis functions and Kalman filter support vector updates for time series prediction. The approach was verified using well-known chaotic time series, but no comparison was made to other SVM prediction techniques. In 2001, Gretton showed that the combination of time-frequency techniques and SVMs outperformed other nonstationary signal classification approaches (e.g., linear discriminant of Wigner distributions, the MCMC method) [81]. Verification was performed using linear chirps. Cohen's class of TFDs was used for analysis; an SVM with a radial basis function kernel was used for classification. Davy and Godsill reported application of a similar technique to unsupervised segmentation of audio signals [82]. In 2002, Chang et al. evaluated competing SVMs for unsupervised time series segmentation by SVMs [83]. The results were comparable to those obtained using the benchmark annealed competition of experts algorithm [84].

Diverse but limited applications of SVMs to time series analysis indicate that they compete favorably with state-of-the-art methods and are less sensitive to researcher parameter selection issues than competing techniques. The success of SVMs for image classification suggests their application to TFD classification as a method of capturing time series dynamics.

1.4.3.2 SVM Classification in Biomedical Engineering

Many of the applications of SVMs to biomedical engineering problems have focused on image processing tasks. Improvements over existing algorithms are often obtained by classifier replacement with an SVM. Some examples include: classification of blood cells by shape, color, and texture features [85]; medical image segmentation under shape deformation [86]; and image segmentation for barcode location [87]. The general problems of parameter specification and kernel choice are usually addressed empirically, but some recent research attempts automatic model specification. In 1999, Chappelle et al. reported a noteworthy example—albeit a non-biomedical engineering example—of image classification that showed the generalization capability of SVMs [88]. Using a commercial collection of digital photos, they computed global color histograms to classify images according to category (e.g., plane, door, bird, helicopter). Despite translation, rotation, scale, and perspective differences they achieved a classification accuracy of 89%. A k-nearest neighbor approach could only achieve 52.3% correct classification.

Liang and Lin describe a straightforward application of SVMs to electrophysiological signal classification [89]. They reported diagnosis of delayed gastric emptying using cutaneous electrogastrogram time series. Data for 152 patients were

windowed (2-minute) and simple frequency domain features (e.g., energy in a band, dominant frequency) were extracted for classification. Their results were superior to previous neural network methods. In an arrhythmia classification application, Strauss et al. successfully used wavelet features extracted from endocardial electrograms [90]. Rojo-Alvarez et al. used SVM classification to discriminate between supraventricular and ventricular tachycardia [91]. The SVM approach improved the interpretability of their results. Millet-Roig et al. showed that SVM classification of malignant arrhythmias outperformed logistic regression and neural network approaches [92]. Features were extracted from 4-second wavelet decompositions of cardiac data. Verification was performed using the AHA database. In 2002, Gorur et al. report improved sleep spindle detection using SVM classification of short-time Fourier transform features [93]. Niyogi et al. reported successful stop consonant identification using spectral features derived from speech time series [94]. The approach outperformed a hidden Markov model operating on cepstral features.

There have been many successful applications of SVMs to biomedical signal and image classification. The central problems of feature representation and time series encoding remain open research issues, but familiar frame-based approaches often work well. Repeated comparisons of SVMs to neural network (and other) approaches generally favor the SVM in terms of computational complexity and accuracy. Strong generalization ability translates into smaller data requirements and increased robustness to parameter selection.

1.4.3.3 Novelty Detection

In 1998, Schölkopf et al. introduced the idea of estimating the support of a distribution using SVMs [95]. Their motivation was to solve a simplified version of the density estimation problem: finding a minimum volume quantile estimator that is "simple." Work on this problem began in 1963 [96], but was mysteriously abandoned. Similar ideas have also appeared in the mathematics literature (e.g., [97]). The solution that Schölkopf et al. arrived at, the 1-class υ-SVM (referred to throughout this work as the 1-class SVM), was inspired by Ben-David's previous work on density estimation in learning [98] and its use for novelty detection was introduced [99].

In the Schölkopf 1-class formulation, data in feature space is separated from the origin by a maximum margin hyperplane. This results in a constrained quadratic programming problem, similar to the basic SVM case, for determining the support vectors [100]. The choice of separating point is arbitrary and affects the decision boundary returned by the algorithm [101]. This is an important but unstudied issue. Basic theoretical properties of the 1-class SVM were proven in the initial and subsequent papers [95, 99, 102]. The most important result is the interpretation of υ previously mentioned. This property is what allows 1-class SVMs to be used as novelty detectors. Schölkopf et al. illustrated this with an application to classifying images from the U. S. Postal Service database of handwritten digits. This database contains 9298 images of size 16 x 16. It is folklore in the machine learning community that the database contains several images that are impossible to classify because of mislabeling or poor segmentation. Schölkopf et al. used the 1-class SVM to rank images in order of novelty

and found that the most novel images were indeed incorrectly labeled or meaningless [99].

An application of the 1-class SVM to fault prognostics of jet engines was reported by Hayton et al. [70]. The motivation behind this research was to develop a fault prognostication technique that could distinguish between "good" and "faulty" engines, a technique that did not require the destruction of expensive engines for training. A novelty detector using the 1-class SVM was trained on frequency-domain features extracted from acoustic time series recorded from normal engines. The novelty detection approach was successful. A cursory investigation of the effect of υ on performance confirmed the interpretation of the parameter and its usefulness for bounding the computational complexity of the algorithm.

Other successful applications of the 1-class SVM have been reported for network intrusion detection [103], personal email (spam) filtering [104], document retrieval [105], and content-based image retrieval (CBIR) from image databases [106]. Methodologies proceed in a similar fashion: domain-specific features are selected (e.g., word frequency in text and spam filtering applications, network usage statistics for intrusion detection, and relevance feedback in CBIR), the classifier is trained and evaluated for different υ- and kernel parameters, and then performance is validated. Training sets tend to be small—on the order of hundreds or fewer instances—but the successes of the applications confirms the generalization capability of SVMs. Some extensions that incorporate partial knowledge of the distribution of negative exemplars have been explored (e.g., [104]) with inconclusive results. A kernel-based version of vector quantization that is similar in spirit

to the 1-class SVM has recently been proposed [107], as well as an alternative approach for novelty detection [108].

As with the basic SVM, there is no automatic method for specifying 1-class SVM model parameters, but the interpretation of υ eases this task to some degree. The use of the 1-class SVM for novelty detection appears promising; however, few rigorously validated applications have been reported. Important, unresearched issues regarding the 1-class SVM include the choice of separating point in feature space and the effect of distribution properties (e.g., skewed, multimodal, heavy-tailed or discontinuous distributions) on the decision surface.

1.5 Contributions

The contributions of this research are:

- Development of a framework for novelty detection using 1-class SVMs that is applicable for analysis of physiological time series like the IEEG.
- Demonstration that the novelty detection framework surpasses the state-of-the-art in seizure analysis as measured by the metrics of sensitivity, mean detection latency, and early-detection fraction
- Implementation of an online novelty detection framework, the naïve online novelty detector, for analysis of time series.
- Demonstration that the naïve online novelty detector also surpasses the state-of-the-art in seizure analysis as measured by the metrics of sensitivity, mean detection latency, and early-detection fraction.

- Demonstration that a novelty detection seizure analysis system can detect seizures using only 15-minute epochs for training with no prior knowledge of a seizure. This eliminates patient-specific tuning of the detector.
- Demonstration of a technique based on the novelty detection framework for mapping the spread of seizure onsets across multiple channels in patient recordings. This technique is useful for automatic focus channel localization.

Chapter 2

Background

2.1 Epilepsy

Epilepsy is the second most prevalent neurological disorder in humans after stroke. It is characterized by recurring seizures in which abnormal electrical activity in the brain causes altered perception or behavior. Approximately 1% of the world population is affected. Documented cases date back to 2000 B.C., but epilepsy was not widely acknowledged as a brain disorder until the 19th century. The mechanisms underlying ictogenesis have been extensively studied by researchers using electrophysiological, pharmacological, and biochemical techniques, but no consensus has been reached on how or why seizures develop. Well-known causes of epilepsy may include: genetic disorders, traumatic brain injury, metabolic disturbances, alcohol or drug abuse, brain tumor, stroke, infection, and cortical malformations (dysplasia) [109]. Current treatments for epilepsy include: antiepileptic drugs, resective surgery, devices such as the vagal nerve stimulator, and the ketogenic diet. In addition, promising new therapies, such as direct electrical brain stimulation, are being developed.

A seizure is a disturbance characterized by changes in neuronal electrochemical activity that results in abnormal synchronous discharges in a large cell population, giving rise to clinical symptoms and signs. The process through which this abnormal synchrony is initiated and cells are recruited and entrained to fire together is called ictogenesis.

Patients experience varied symptoms during seizures depending on the location and extent of the affected brain tissue. Symptoms may include involuntary clonic movements, an altered state of awareness, convulsions, or impairment of consciousness, unusual or repetitive behaviors, or odd sensations. Seizures generally last from a few seconds to a few minutes. Seizures can be classified in several ways based upon the site and extent of the brain that is affected, clinical symptoms, EEG pattern, and etiology [110]. Partial seizures are limited in extent; generalized seizures typically affect the entire brain and impair consciousness. A taxonomy of seizures is shown in Figure 2-1. The adult patients in the available database exhibit complex partial seizures originating focally in the temporal lobes.

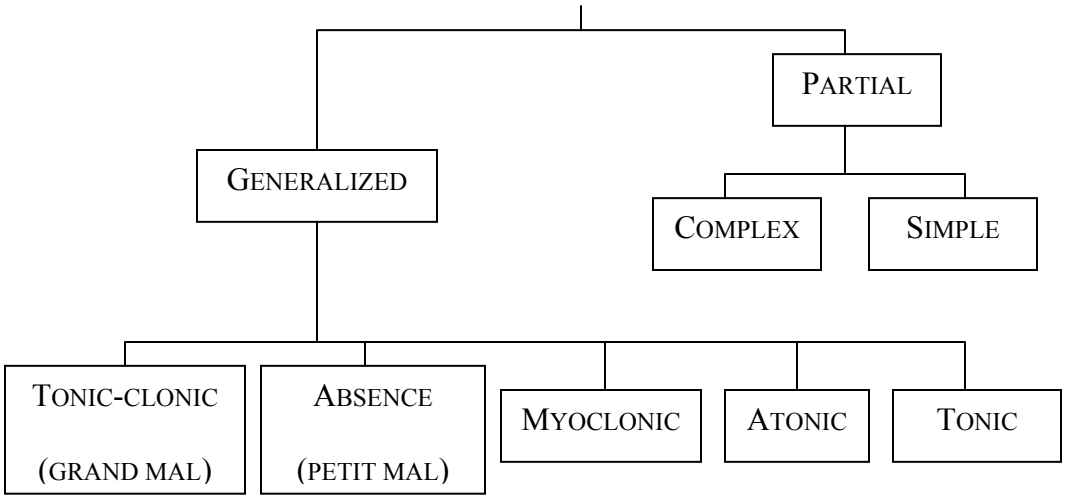

Fig. 2-1. A taxonomy of seizures based upon classifications from the International League Against Epilepsy (1985).

2.2 The Electroencephalogram (EEG)

The first known recording of an EEG was made in 1929. Many EEG recording techniques are available; they differ by recording location or type of electrode used for measurement (cf. 1.2). Scalp EEG is the least invasive type of EEG, but it provides a distorted, lowpass-filtered signal because of the effects of intervening skull, tissue, and cerebrospinal fluid. The effective imaging depth attainable using scalp EEG is approximately 1 cm. Scalp EEG is also severely affected by motion, muscle, and recording artifacts. Intracranial EEGs provide high signal-to-noise ratio measurements that are relatively artifact free. They are recorded by placing subdural grids or strips on the cerebral cortex, or inserting depth electrodes into the mesial temporal lobe and other subcortical regions. The ability to collect IEEG data is limited to clinical necessity during presurgical evaluation because of ethical issues and the highly invasive nature of the technique. Other complementary functional imaging techniques are available (e.g. functional magnetic resonance imaging (fMRI), positron emission tomography (PET)), and single photon emission computed tomography (SPECT), but the EEG is the primary tool used in diagnosing epilepsy.

EEG signals are commonly analyzed by identifying morphological features, (e.g. spike-and-slow-waves, beta buzzes, bursts, etc.) or examining frequency bands associated with different mental activity or conscious state. Table 2-1 shows these common frequency bands.

Table 2-1. Commonly examined frequency bands of interest in EEG analysis.

Name	Frequency Range (Hz)	Associations
Delta	0 – 4	Deep sleep, serious brain disease or anesthesia
Theta	4 – 8	Emotional stress; drowsiness (abnormal in wakefulness)
Alpha	8 – 13	Quiet wakeful state (eyes closed)
Beta	13 – 30	Intense mental activity, stress and tension
Gamma	30 +	Observed at the cellular level; previously not of interest

2.3 The Epilepsy Database

2.3.1 Overview

The corpus of data available for this research consists of a database containing IEEG recordings of epilepsy patients made during evaluation for epilepsy surgery with intracranial electrodes. Patients were diagnosed with mesial temporal lobe epilepsy where lateralization to one or both temporal lobes was not possible or definitive during scalp video-EEG monitoring. Patients were observed for 3 to 14 days, and between 8 and 275 hours of data were recorded. Data quantity and quality vary by patient. Some recordings have large gaps where no data is available due to patients leaving monitoring for brain imaging or other testing, patients removing electrodes during seizures, or

equipment malfunctions. Data were collected for 18 patients using a Nicolet 5000 video EEG acquisition system. Between 20 and 36 electrodes were placed intracranially, and simultaneous EEG and video recordings were made. Figure 2-2 shows the electrode locations.

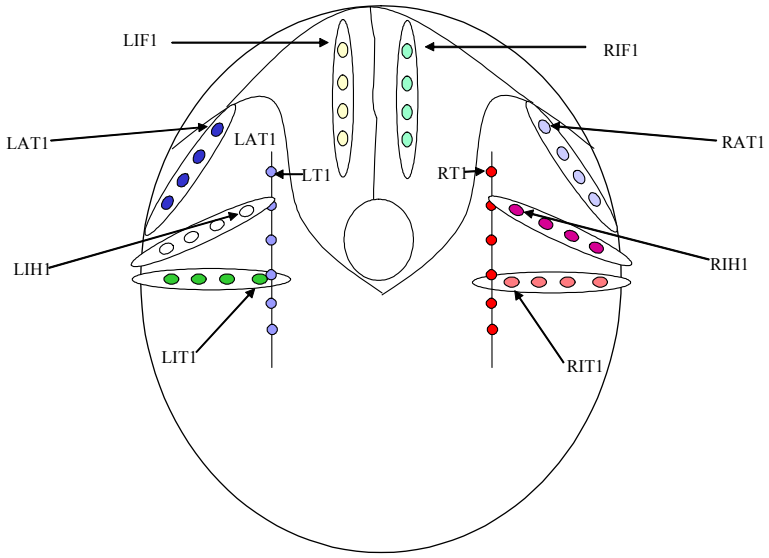

Fig 2-2. Electrode placement for patients in the corpus. LT: left temporal lobe, RT: right temporal lobe, LIF: left inferior frontal lobe, RIF: right inferior frontal lobe, LIT: left inferior temporal lobe, RIT: right inferior temporal lobe, LAT: left anterior temporal lobe, RAT: right anterior temporal lobe, LIH: left hippocampal formation, RIH: right hippocampal formation. Note that the most distal contact, or the electrode contact closest to the probe tip, is referred to as contact #1.

Figure 2-3 shows the placement of a depth electrode along the axis of the hippocampus via the "occipital approach," in the left temporal lobe of one study patient.

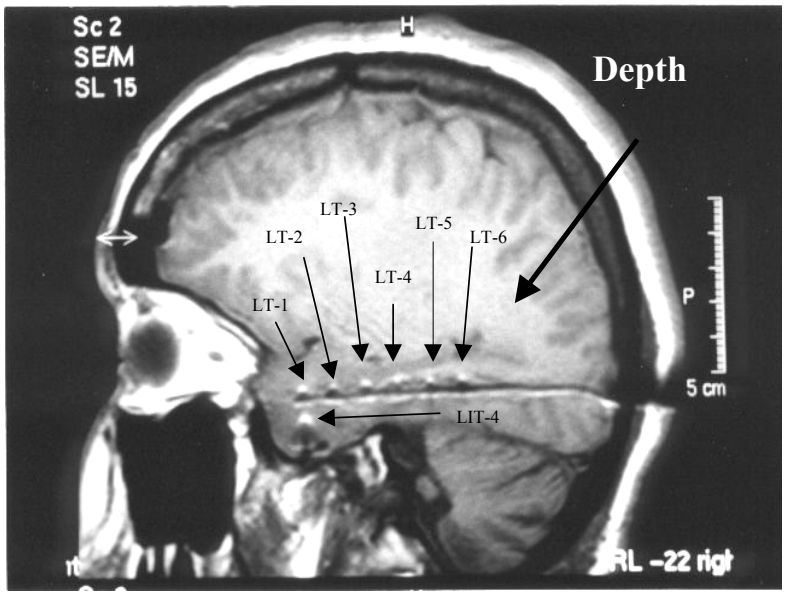

Fig 2-3. MRI showing depth electrode placement in a patient.

The EEG data were amplified, bandpass filtered (0.1 Hz and 100 Hz cutoffs), and digitized at 200 Hz (12 bits/sample resolution). More than 1077 hours of adult patient data are available. Table 2-2 summarizes details of the data by patient. More extensive details on this database are available [2, 4]. All data in the corpus have been reviewed by neurologists and marked to indicate sleep state, seizure onsets, and recording artifacts.

Table 2-2. Summary of dataset selected for analysis, including patient number, number of recording epochs per patient, total duration (*hh:mm:ss*) of recordings per patient, number of seizures observed for each patient, and the primary focus channel.

Patient	Epochs	Duration	Seizures	Focus Channel
1	6	46:04:39	6	LT2
2	6	47:03:22	5	RIT4
3	25	189:03:27	10	RT5
4	14	99:01:06	10	RT2
5	5	38:42:55	5	RT2
7	8	57:25:13	11	LT2
8	20	151:10:41	4	LT2
9	20	156:03:23	11	RT2
10	8	49:59:45	13	LT3
11	5	39:20:00	5	RT2
15	8	58:12:41	7	LT2
16	18	145:34:24	31	LT3
All	*143*	*1077:41:36*	*118*	–

2.3.2 Patient Annotations

The details of the epilepsies exhibited vary by patient. These details are relevant for selecting patient data for analysis, and for noting artifacts or other phenomena relevant to performance assessment. Annotations for patients used in this research follow.

Patient 1 exhibited six stereotyped, unilateral, well-localized focal seizures originating near the head of the left hippocampus at electrode locations LT2 and LT3. All seizures were preceded by 20-30 Hz oscillations, with "chirp" morphology.

Patient 2 exhibited five stereotyped, unilateral, well-localized focal seizures originating at electrode location RIT4. All seizures were preceded by a long beta buzz.

Patient 3 exhibited ten seizures originating from the right hippocampus near electrode locations RT5 and RT6. Most seizures were preceded by a generalized flattening and loss of background. Some data were corrupted by reference artifacts.

Patient 4 exhibited ten stereotyped, unilateral, well-localized focal seizures. Eight of the seizures originated near electrode locations RT2, RT3, and RT4, all in the right anterior to mid hippocampus; two of the seizures originated near electrode locations LT3, in the left anterior to mid hippocampus. The last six seizures occurred in a flurry during a 10-hour recording period.

Patient 5 exhibited five stereotyped, unilateral, well-localized focal seizures originating consistently near electrode locations RT2 and RT3 (right anterior to mid hippocampus).

Patient 7 exhibited seven focal seizures originating primarily in the left temporal lobe. Epileptologist notes indicate that the seizures are accompanied by diffuse decrement (decreased amplitude), followed by focal fast activity, suggesting that the electrodes may not be optimally placed in this patient, or that there is rapid spread of focal activity to the thalamus early in seizure generation.

Patient 8 exhibited four focal seizures originating from both right and left temporal lobes near electrode locations LT1 and RT4.

Patient 9 exhibited eleven focal seizures originating on the right side of the brain between the right temporal lobe and the right inferior frontal region near electrode locations RT, RT5, and RIF2.

Patient 10 exhibited thirteen focal seizures originating in the left temporal lobe at electrode locations LT3 and LT4 (mid-hippocampus). Epileptologist notes indicate the many seizures show precursor activity in the right temporal lobe (e.g. delta slowing), and that many records are badly distorted by reference artifacts.

Patient 11 exhibited five focal seizures originating in the right temporal lobe near electrode locations RT2 and RIT2. Two of the recording epochs were corrupted by reference artifacts and distortion.

Patient 15 exhibited seven focal seizures originating from both the left and right temporal lobes near electrode locations LT2 and RT2. Electrode RT2 is high impedance and contains intermittent, high-frequency artifacts.

Patient 16 exhibited thirteen focal seizures originating in the left temporal lobe at electrode locations LT2, LT3, and LT4. Six of the seizures were only marked by the epileptologist after an automated seizure detection algorithm indicated epileptiform activity was present and previously unnoted.

It is important to note here that all patient clinical information, marking of seizure onsets and verification of clinical information was performed independently by two academic epilepsy specialists, each boarded in Neurology and Clinical Neurophysiology.

Chapter 3

Research Methodology

A block diagram of the key stages of this research is shown in Figure 3-1. The methodology followed a traditional machine learning approach: (1) data were collected and labeled; (2) the data were appropriately preprocessed; (3) features were selected and extracted; (4) a classifier was trained; and (5) the performance of the system is evaluated. The entire process was iterated until acceptable performance results were achieved. The research was conducted in two phases: offline algorithm development, and online algorithm development. The results from the former influenced the design of the latter. Learning is applicable only to the online algorithm.

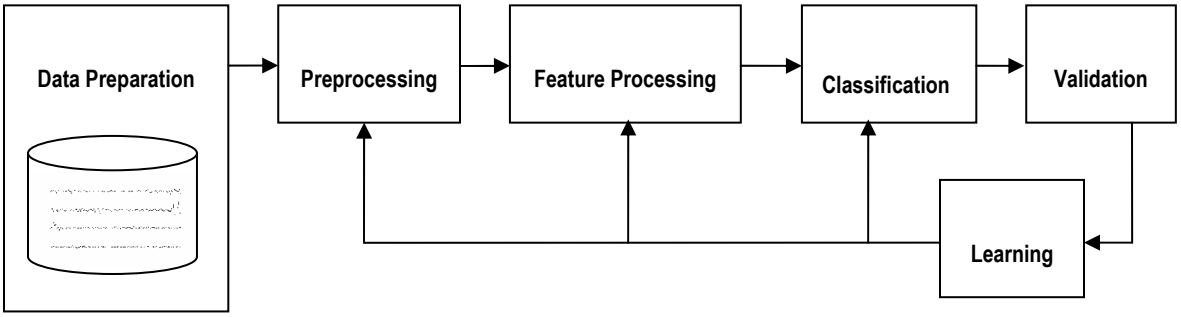

Fig. 3-1. An overview of the research methodology.

The following sections describe each of the major steps in greater detail. The following chapters then presents key results and discussion of performance.

3.1 Data Preparation

3.1.1 Data Warehousing

Data from an epilepsy corpus was available for analysis (cf. 2.3). The data consist of a patient-specific number of multichannel IEEG recordings, typically 28 channels of recordings, and simultaneous video recordings; in addition, ECG and other EEG data (e.g., scalp electrodes for sleep staging) were collected for some patients. Only the IEEG data were considered in this work. The electrographic data were bandpass filtered (frequency cutoffs at 0.1 Hz and 100 Hz) and then sampled at 200 Hz (12 bits/sample at 1.08507 μV/bit resolution). Metadata accompanying each recording was available including patient number, focus channel location(s), UEO and EEC times, sleep state information, and drug dosing schedule. The data and metadata were not collated and accessible in a convenient form for analysis. One goal of this research was to design and populate a data warehouse capable of supporting semi-automated analysis of the data. This entailed database design, data conversion, data cleansing, and the development of convenient access and maintenance tools. A popular open source database, mySQL, was used for storage and management of metadata. A custom MATLAB toolbox was developed to support semi-automated data analysis tasks.

3.1.2 Patient and Channel Selection

Most research on seizure analysis has concentrated on focal seizures and relied on *a priori* knowledge of the focus channel. In this research, patient data was grouped according to the following criteria for further analysis:

(1) *Canonical focal data*—data originating from patients implanted with intracranial electrodes that clearly exhibit well-localized, stereotyped focal seizures. Patients 1, 4, 5, 7, and 10 recordings satisfied these criteria.

(2) *Qualified data*—the remaining data in the corpora (e.g., Patients 2, 3, 8, 9, 11, 15, and 16 recordings).

While it is desirable to make full use of the available data, the development of the offline detector was performed on only the canonical focal dataset (i.e., the channel(s) in or nearest the epileptic focus). This decision was made to facilitate algorithm development and characterization. Both the offline and online detectors performed single-channel analysis. No attempt was made to incorporate multichannel features. Multichannel analysis was performed during experiments dedicated to mapping the location of the seizure onset zone.

3.1.3 Baseline and Onset Epoch Selection

An advantage of the novelty detection framework for seizure analysis is that it requires only baseline data for training—no seizure data is necessary for detector training. Seizure epochs are required for performance assessment, however. Figure 3-2

shows typical single-channel baseline and onset epochs used in the development of the offline detector, and for training the 1-class SVM in the online detector.

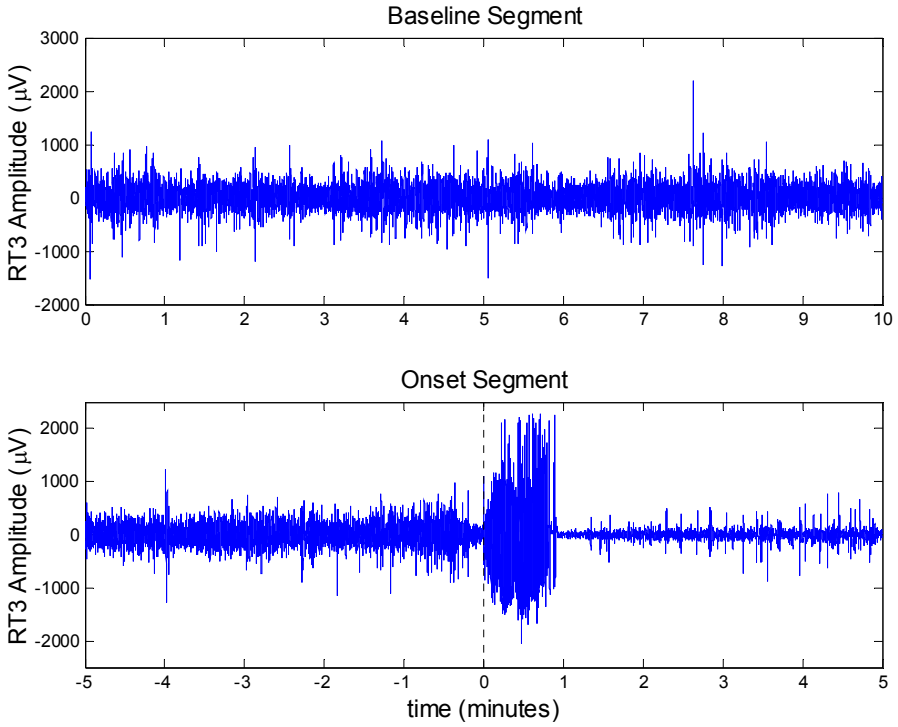

Fig. 3-2. Typical baseline and onset epochs of IEEG data measured at electrode RT3; note the relative location of the UEO.

3.2 Preprocessing

3.2.1 Sleep Staging

Knowledge of patient sleep state is crucial to the correct interpretation of EEG activity. The corpora do not contain polysomnogram data, so clinical sleep staging by

traditional means is not possible. Expert analyses of video recordings made during patient observations provide a gross indicator of sleep state: sleepy, drowsy or awake. However, they are too imprecise and error-prone to incorporate as sleep staging features.

For patients where the ECG was recorded, it may be possible to accurately determine sleep state by analyzing the *heart rate variability (HRV)* signal [111, 112]. The HRV signal is defined as the sequence of interval times occurring between successive heart beats. It is typically derived from the ECG by automatic measurement of the cardiac R-R interval time. In normal human subjects it has been shown that HRV indicates changes in patient sleep state minutes before the EEG power spectra [113, 114]. It is not clear that these results hold for patients with epilepsy [36-39], however, this approach may merit further investigation.

No explicit sleep staging features were available for use in the IEEG analysis by the static detector. It is hypothesized that this contributed significantly to the mean false positive rate of the detector, given the known sensitivity of seizure prediction methods to patient state of awareness [50].

3.2.2 Denoising and Artifact Removal

The recorded data exhibit high signal-to-noise-ratio (SNR) but are still corrupted by noise. The primary sources of systematic noise are power line coupling and electrode motion. Several techniques can be employed to minimize the effect of line noise. These include (in increasing order of effectiveness): bipolar montaging, notch filtering, adaptive filtering and independent component analysis (ICA) [115, 116]. In bipolar montaging two spatially adjacent channels are combined via subtraction to eliminate common-mode signals. The simple subtraction of two signals, however, does not

account for the covariance structure of the data and may eliminate common-mode cortically generated signals of importance. In a preliminary analysis of the data it was observed that bipolar montaging decreased the variance of the resultant signal and improved the overall classification performance. A notch- or comb filter is an infinite impulse response (IIR) filter for zeroing specific frequency contributions (e.g. 60 Hz). Adaptive filtering and ICA are not implemented in this research because of their computational complexity.

In addition to line noise, artifacts due to electrode movement are common in electrophysiological recordings. The data were visually inspected to identify segments containing artifacts; this information was stored in the data warehouse. Artifact contaminated segments were excluded from analysis in this work.

3.3 Feature Processing

Feature processing refers here to three tasks: feature selection, transformation, and extraction. The selection and extraction of "good" features, an important machine learning topic in its own right, directly influences the performance of a classification system in terms of accuracy, required learning time, necessary exemplars and computational complexity.

3.3.1 Feature Selection

The features used in this research were selected after a literature review. These features are strong measures of signal energy, and do not explicitly account for the evolution of time-frequency signal dynamics. However, these features have been

reported in successful seizure detection algorithms. Collectively, they are referred to in this work as *benchmark features*.

3.3.1.1 Benchmark Features

Many features have been proposed for seizure analysis (see section 1.4). Three features have proven especially effective for seizure detection [2, 4, 28, 30]: mean curve length, $CL[n]$; mean energy, $E[n]$; and mean Teager energy, $TE[n]$. These are primarily energy-related features computed over short data windows. The *benchmark feature vector* used in this research consists of logarithmically-scaled variants of these features given by:

$$CL'[n] = \log\left(\frac{1}{N}\sum_{m=n-N+1}^{n}|x[m]-x[m-1]|\right) \tag{3.1}$$

$$E'[n] = \log\left(\frac{1}{N}\sum_{m=n-N+1}^{n}x[m]^2\right) \tag{3.2}$$

$$TE'[n] = \log\left(\frac{1}{N}\sum_{m=n-N+3}^{n}\left(x[m-1]^2 - x[m]x[m-2]\right)\right) \tag{3.3}$$

where $x[m]$ is a physiological time series and N is the window length. The selection of window length is an important issue discussed in Section 3.3.2.

3.3.1.2 Feature Space Efficiency

The efficiency of features is often a consideration in system design. Using an uncorrelated set of features can improve the efficiency of a classifier, and the proper selection of features directly affects classifier accuracy. During early experiments it was observed that the feature set considered in this research showed some correlations. Calculation of the Pearson correlation coefficient revealed that two of the features, the average Teager energy (TE) and the curve length (CL), were correlated at the 0.95 level for all patient data. Figure 3-3 shows 2-D slices of the feature space, along with the quantile estimation contour from the 1-class SVM; the strong linear relationship between CL and TE is evident. Also note that the feature space is unimodal and heavy-tailed, so it is easily modeled by the SVM.

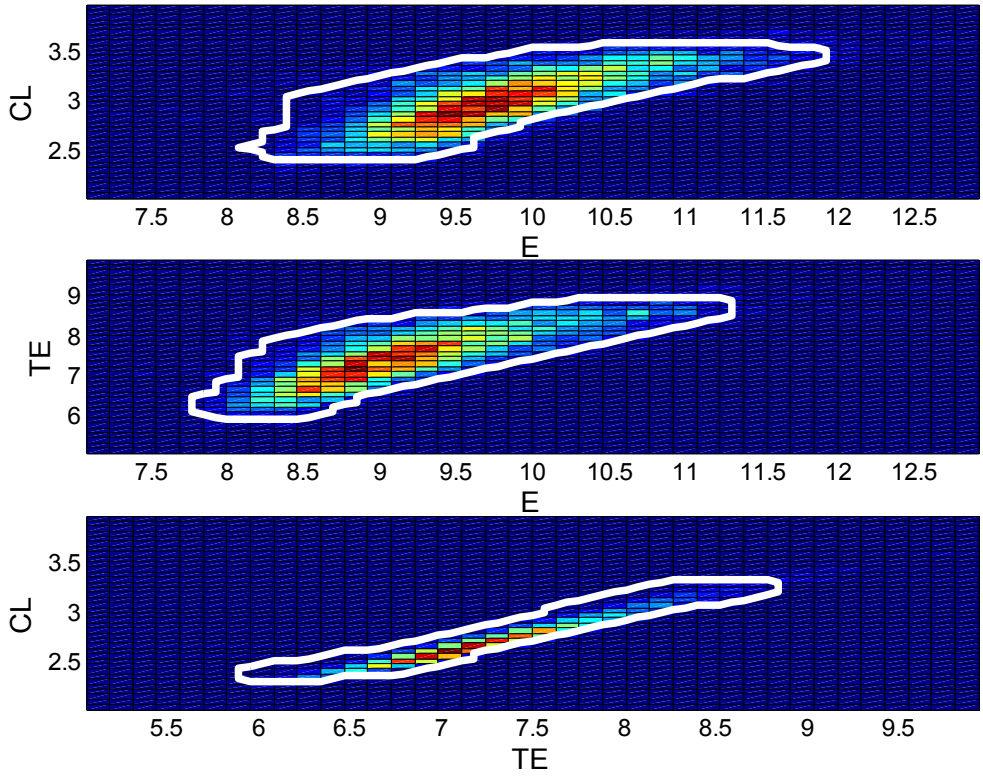

Fig. 3-3. Slices of the novelty detector feature space: CL (curve length), TE (Teager energy), and E (energy). Also shown are the 1-class contours corresponding to the 90-percentile estimate. These data were taken from Patient 5 baselines, but are representative of the entire data set.

3.3.2 Feature Transformation

In practice, a transformation from input space to feature space, $x \mapsto \Phi(x)$, may be performed by a composition of functions, $\Phi(x) = f_1 \circ \cdots \circ f_n(x)$, as in the mean curve length feature defined in Section 2.4. This compositional construction is referred to as *feature transformation*. Sometimes the resulting representation in feature space is referred to as a *hierarchical feature*. It is noted that some researchers distinguish between "raw" and hierarchical features; no such distinction is made in this work.

3.3.3 Feature Extraction

Features are extracted using a block processing approach. In block processing, the data are windowed, a feature vector is computed, and the window is advanced in time. The length of the analysis window is an important parameter in block processing schemes. Many researchers have investigated optimal window lengths for EEG analysis under the short-time stationarity hypothesis with results ranging from 0.25-seconds – 5-seconds [2, 117]. In this research the default window size is fixed at 1.0-seconds. Window overlap is chosen to be 50% of the window length which corresponds to an overlap time of 0.5-seconds.

3.3.4 Feature Normalization

Feature vectors may need to be scaled, or normalized prior to classifier training. It is common practice among SVM researchers to scale all features vectors to the range [0, 1], or alternatively to the range [-1, 1]. The former scaling method is adopted in this research.

3.4 System Architecture

Seizure analysis results have been reported for many different classification architectures (e.g. [2, 4, 22, 57, 118]). The SVM is a popular alternative to traditional and neural network classifiers with compelling empirical and theoretical properties. Several formulations and extensions to the classic SVM have been developed [100]. One formulation, the 1-class SVM, is suitable for novelty detection applications [99] and was utilized in this research. The classification architecture employed for applying novelty detection to seizure analysis is shown in Figure 3-4.

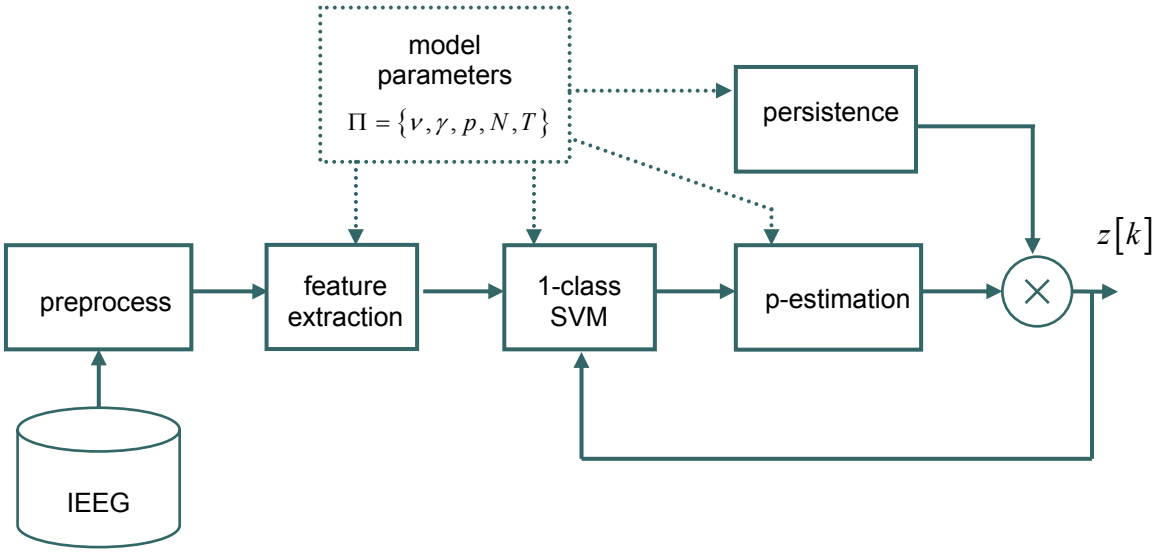

Fig. 3-4. The architecture for the novelty detection framework for seizure analysis. IEEG time series data is processed in stages to produce the final output, $z[n]$, which is the frame-wise classification sequence of IEEG as *novel* or *normal*.

IEEG data are processed in stages to produce a label for each frame of data. For adaptive algorithms (e.g., the naïve online detector to be presented in Chapter 5), the output updates the 1-class SVM according to the adaptation strategy employed. The novelty detection problem and ν-SVM classifier are briefly reviewed in the following sections, along with a simple approach to hypothesis testing.

3.4.1 Novelty Detection

Previous work has posed seizure analysis as a binary classification problem, e.g. "seizure" versus "no-seizure", "preseizure" versus "no-preseizure." This formulation suffers from the following drawbacks:

1. Training a binary classifier capable of strong generalization is difficult if there is a large imbalance between the number of positive and negative examples presented for training. Such situations are representative of *class imbalance problems* addressed in machine learning research. This is relevant to seizure analysis since the fraction of "seizure" data is estimated at less than 0.05.
2. Two-class approaches assume that both positive and negative examples cluster in specific ways. It is not clear that this is true (consider the question, "how many measurably distinct preictal states does the EEG exhibit?"). What is more evident, is that interesting events in seizure analysis are infrequent and differ from normal data. A one-class-with-outliers viewpoint is more intuitive.
3. The two-class approach implies a characteristic time horizon for prediction. For example, when considering "preseizure" versus "seizure" segments, do "preseizure" data begin 1-minute prior to onset? 5-minutes? 20-minutes? Are all seizures preceded by stereotypical precursors beginning precisely 1- or 5- or 20-minutes

prior to onset, or is it possible that these precursors are stochastically distributed?

This research approached the seizure analysis problem from a 1-class perspective. In 1-class classification, it is only assumed that positive examples cluster in a consistent manner. Outliers amongst the positive examples (novelties) are classified as negative examples. Such an approach is referred to as novelty detection. Figure 3-5 illustrates the difference between traditional models of brain electrical activity versus the novelty model.

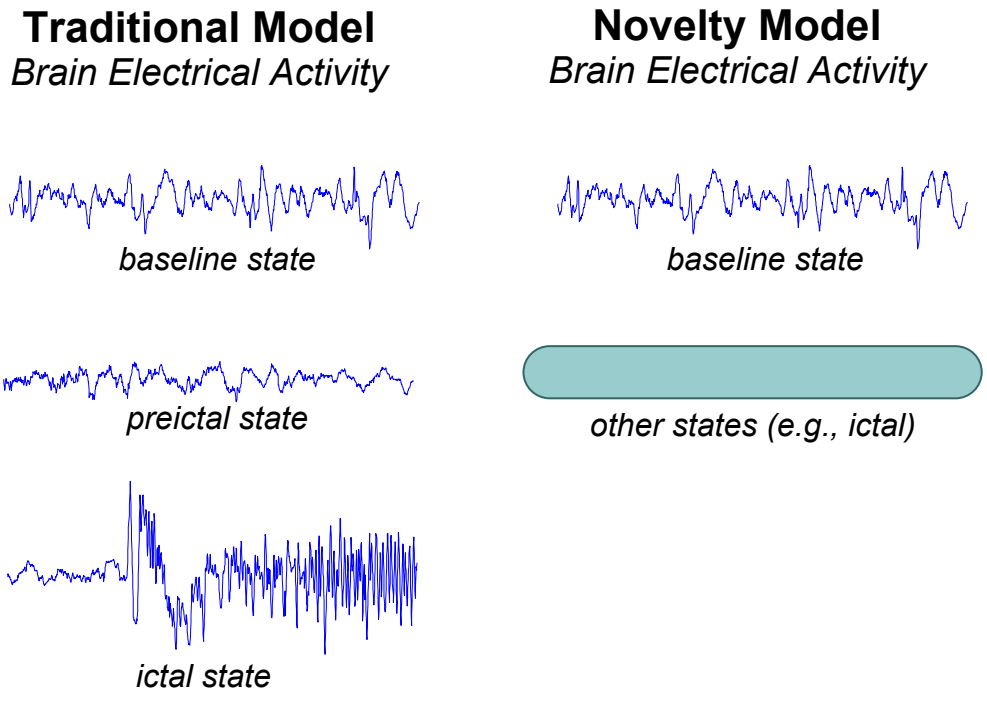

Fig. 3-5. Brain electrical activity models imposed by epilepsy researchers—traditional models assume *a priori* known classes.

A more precise definition of novelty detection is stated in Definition 3.1.

Definition 3.1 (Novelty Detection). *Given a set of iid training samples,* $x_1, \cdots, x_n \in X \subseteq \Re^N$*, drawn from a probability distribution in feature space,* P*, the goal of novelty detection is to determine the "simplest" subset,* S*, of the feature space such that the probability that an unseen test point,* x'*, drawn from* P *lies outside of* S *is bounded by an a priori specified value,* $\upsilon \in (0,1]$.

This research adopted a novelty detection approach for classifying IEEG time series data. Baseline brain electrical activity was regarded as the positive, or *normal*, class. It was hypothesized that feature vectors corresponding to artifact- or seizure-related frames were distributed differently in feature space, and that seizure events would be detected as novelties. This assumption is dependent on the true nature of ictogenesis, the discriminative power of the features selected, and the technique adopted for novelty detection.

3.4.2 1-class SVM

The 1-class SVM is one approach to solving the novelty detection problem [100]. In this formulation, the data is first mapped into a feature space using an appropriate kernel function, and then it is maximally separated from the origin using a hyperplane. The hyperplane parameters are determined by solving a quadratic programming problem:

$$\min\left(\frac{1}{2}\|w\|^2 + \frac{1}{\upsilon l}\sum_{i=1}^{l}\xi_i - \rho\right) \tag{3.4}$$

$$\text{subject to } \left(w \cdot \Phi(x_i)\right) \geq \rho - \xi_i \quad i = 1,2,\ldots,l \quad \xi_i \geq 0 \tag{3.5}$$

where w and ρ are hyperplane parameters, Φ is the map from the input space to feature space, υ is the asymptotic fraction of outliers (novelties) allowed, l is the number of training instances and ξ is a slack variable. Figure 3-6 shows the geometry of the single-class SVM in feature space.

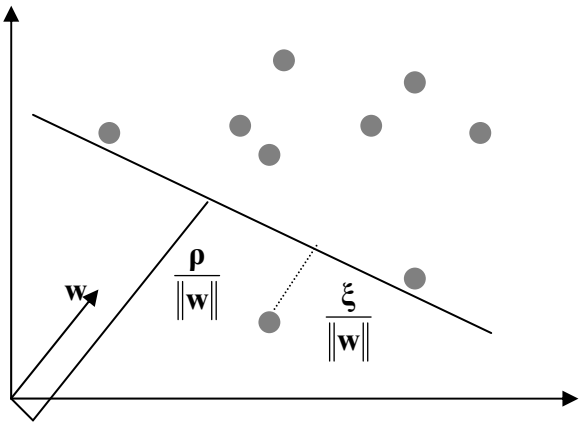

Fig. 3-6. Geometry of the 1-class SVM in feature space. Note the hyperplane and associated parameters, ρ and w , and the slack-variable, ξ , penalizing misclassifications.

The implementation of the 1-class SVM algorithm requires the following model parameter specifications: kernel function, kernel parameters, outlier fraction, and

separating point in feature space. The kernel function encodes prior information about the problem (e.g., distance or similarity measures). The most common kernel function used is the radial basis function which is described by a single parameter, the kernel width, γ. The outlier fraction should incorporate prior knowledge regarding the frequency of novelty occurrences (i.e., seizing frequency); a nominal value used in this research is $\nu = 0.1$, indicating that approximately 10% or less of the entire data is estimated to be novel. The smaller the value of ν, the more efficient the SVM computation becomes, and the smaller the fraction of data that is declared novel. The choice of separating point in feature space and its effects on the decision boundary is an open research issue. In this work, the Schölkopf formulation—separation in feature space from the origin—was implemented. An important task in this research was the characterization of parameter effects on detector performance.

An integrated library of SVM tools, LIBSVM, is freely-available and was used for the υ-SVM classification tasks. Additional MATLAB tools were developed to support the novelty detection framework.

3.4.3 Maximum Likelihood Event Detection

A fundamental property of the 1-class SVM guarantees that the fraction of instances classified as outliers will asymptotically equal υ for a stationary process. At first glance this suggests that υ determines the specificity of the system in a direct manner (i.e., by setting the constant false alarm rate). This is not the case, however, for two reasons: (1) the asymptotic nature of the υ bound, and (2) the non-stationarity of the physiological signals. The first limitation is generally insignificant in practice; the second is the more serious consideration. While the short-time stationarity assumption

may hold over single analysis frames, it does not imply that the resulting sequence of feature vectors is stationary. This fact is the principle behind seizure analysis by the 1-class SVM: changes in the distribution of feature vectors corresponding to seizure-related events will be detected as novelties. Unfortunately, it is also probably true that changes in the distribution of feature vectors during normal brain electrical activity will occur for unknown reasons, and these changes will also be detected as novelties. The latter adversely affect classification accuracy.

A simple sequential hypothesis test is proposed to control the rate of false alarms. This test is simply an estimate of the probability that IEEG exhibits novel behavior for a frame. For a sequence of N observations of the 1-class SVM output, $y \in \{-1, +1\}$, the following hold:

$$N_+ + N_- = N \tag{3.6}$$

$$N_+ - N_- = \sum_{i=1}^{N} y[i] \tag{3.7}$$

where N_+ and N_- are the number of positive and negative output occurrences, respectively. Under the assumption that outputs are identically independently distributed (*iid*), the maximum likelihood estimate of the probability that a novelty event occurred, $\hat{p}$, is given by:

$$\hat{p} = P\left(novelty \mid y_1, y_{2,} \cdots, y_N\right) \tag{3.8}$$

$$\hat{p} = \frac{N_{-}}{N} = \frac{1}{2}\left(1 - \frac{1}{N}\sum_{i=1}^{N} y[i]\right) \tag{3.9}$$

As discussed, $\lim_{N\to\infty} \hat{p} = \upsilon + \varepsilon$, where $\varepsilon > 0$ is an error term arising from nonstationarity. A decision function for declaring frames as baseline or novelty can be constructed by thresholding:

$$w[k] = \text{sgn}\left(\frac{1}{2}\left(1 - \frac{1}{N}\sum_{i=k-N+1}^{k} y[i]\right) - p\right) \tag{3.10}$$

where $w[k]$ is the k-th frame label (novelty = "−1" or baseline = "+1"), $y[i]$ is the i-th frame classifier output, N is the number of classifier outputs considered and p is a user-specified threshold. The selection of the parameters N and p was a topic for further investigation, and is discussed in more detail in Chapter 4.

3.4.4 Persistence (Detector Refractory Period)

The basic novelty detector operates by classifying individual frames as novel or normal: if a 1-second frame exceeds the novelty detection threshold, p, that frame is declared as novel $\left(z[n] = -1\right)$. It was observed during early experiments that the detector tends to generate novelty events (i.e. "fire") in bursts. Considering novelty events on a frame-wise basis can produce misleading results because of this bursty behavior. For example, multiple detections of a single seizure, or multiple false positive declarations

may occur during a short interval of time. To address this problem, a refractory parameter, T, (also known as persistence) was introduced to the detection system. The refractory parameter specifies an interval during which the detector, if triggered, maintains its state and ignores subsequent triggers. For example, a refractory period of 30 seconds restricts the detector from firing more than a single time during any 30-second interval, regardless of the number of novelty events generated. Figure 3-7 illustrates two examples of system improvement by implementing persistence.

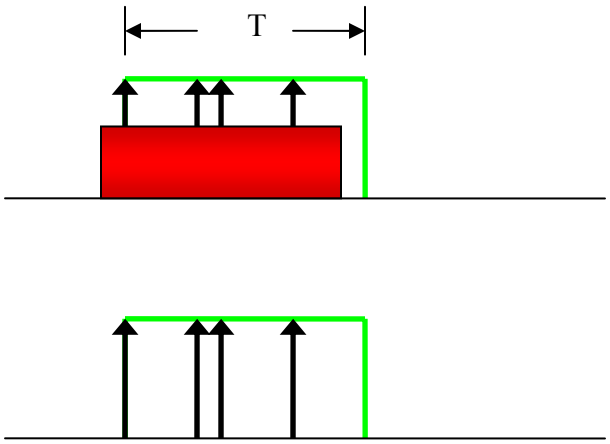

Fig. 3-7. Examples of bursty detector behavior and persistence (*green*). (*Top*) Multiple detections of the same seizure (*red*). (*Bottom*) multiple false positive events are shown. Introducing persistence improves system performance.

The introduction of persistence offers an improvement to the basic system beyond false positive rate improvement: it allows for the characterization of the detector over a range of detection horizons. As the detection horizon (persistence) decreases, one

expects the false positive rate to increase and the detection latency to approach zero seconds. As the detection horizon increases, one expects the false positive rate to decrease, asymptotically approaching a value determined jointly by the novelty parameters of the system (some fraction of the data will always be novel) and the actual novelty rate due to epileptiform activity. The effect of persistence on detector performance was a research issue that is discussed in more detail in Chapter 4.

3.5 Validation

Validation refers to the quantitative assessment of an algorithm's overall performance. Cross-validation was performed for static detector experiments to characterize the performance of the detector.

3.5.1 Cross-validation

With a few notable exceptions (e.g. [2, 4, 28, 31, 119]), most published results on seizure detection and prediction do not demonstrate significant validation of results. It is common practice in machine learning and data mining to perform *k-fold cross-validation* to assess the performance of a classification algorithm [120]. In k-fold cross-validation the data is partitioned into k subsets of approximately equal size. Training and testing is performed k times. During each iteration, a single subset of data is selected for testing and held-out; the other $k-1$ subsets are used for training. When k is equal to the sample size the technique is called *leave-one-out cross-validation (LOO-CV)*. LOO-CV was used to estimate performance metrics for the static detector.

3.5.2 Performance Metrics

Four key performance metrics can be derived from the evaluation of the detector output: sensitivity, early-detection fraction, mean detection latency, and false positive rate. The detector's *sensitivity, (S)*, is a measure of the classification accuracy. It is computed as:

$$S = \frac{TP}{TP + FN} \tag{3.11}$$

where *TP* and *FN* are the number of block true positives and block false negatives (cf. 1.2). The *early-detection fraction, (ED)*, is a measure of the ability of the detector to predict, at least on a short-time basis, an impending seizure onset. It is computed as:

$$ED = \frac{TP\big|_{\tau<0}}{TP\big|_{\tau<0} + TP\big|_{\tau\geq 0}} \tag{3.12}$$

where true positives are discriminated by their detection latency. The *mean detection latency, (μ_τ)*, is computed as:

$$\mu_\tau = \frac{1}{N}\sum_{i=1}^{N} \tau_i \tag{3.13}$$

where τ_i is the detection latency of each detected seizure. Finally, the *false positive rate, (FPR)*, is computed as:

$$FPR = \frac{FP}{T} \tag{3.14}$$

where T is the duration of data (in hours) analyzed. The false positive rate is reported as *false positives per hour (Fph*). It should be noted that FPR can be misleading since different patients generally experience different seizure frequencies.

3.6 Learning

For an online detector, learning is essential. In this research, learning is regarded as a scheduling problem for retraining, or updating, the 1-class SVM.

Chapter 4

Static Detector Results

This chapter presents the results of offline experiments conducted using the novelty detection system as a static seizure detection tool. No adaptation or learning mechanism was incorporated: the detector was trained on epochs of baseline data, and then presented with additional epochs (baseline and seizure) for analysis. Details of the experimental method are presented, including a discussion of data preparation techniques selection criteria, and performance assessment issues associated with offline analysis. Key results are then presented and discussed, with an emphasis on four key performance criteria (c.f. 3.5.2): sensitivity, mean detection latency, early-detection fraction (prediction of seizure onset on EEG), and false positive rate. These findings motivated further experiments to investigate the relationship between detector persistence, detector parameter settings, and system performance. The results of these investigations are presented; they suggest that the system is comparable to state-of-the-art seizure detectors.

4.1 Multi-Patient Focus Channel Experiments

Verification of the novelty detection approach was performed using a subset of the "best" data available in the corpora: this data subset consisted of 81 15-minute epochs of baseline- and ictal IEEG focus channel recordings from five patients. The criteria for evaluating the results of these experiments (in order of priority) were: low false negative rate (FNR), low detection latency, large early-detection fraction, and low

FPR. Results are presented and discussed: with the exception of FPR, the results are comparable to (or exceed) the state-of-the-art. An interpretation of the FPR results is provided, along with an argument for its relative unimportance as a figure of merit that is consistent with a growing trend by epilepsy researchers to pursue hyperdetection strategies for seizure analysis.

4.1.1 Methods

4.1.1.1 Data Preparation

IEEG data from an initial subset of patients were selected for analysis. Recordings from the first five patients with temporal onset seizures requiring video-EEG monitoring with intracranial electrodes were selected: Patients 1, 4, 5, 7, and 10. These data were also considered by Litt et al. [5] in the context of seizure prediction.

The duration of the total available data for the five patients considered is 275 hours. Figure 4-1 depicts the data availability. As noted in Section 2.3, data gaps occurred during patient recording. Current archiving methods employ complete digital recording and archiving, eliminating all data gaps, except for those related to the patient being disconnected from the recording system.

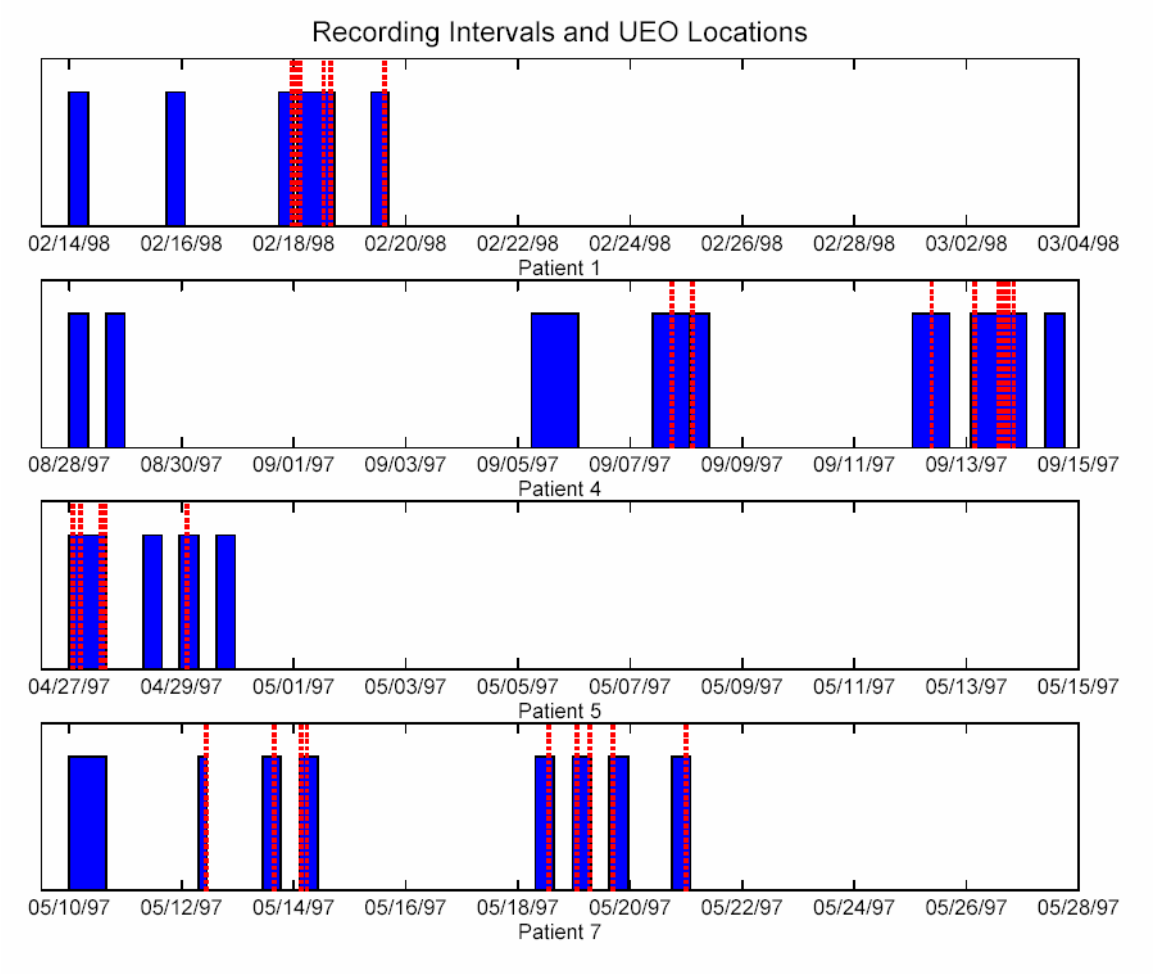

Fig. 4-1. Patient data availability map showing observation periods and available data. Patient 10 is not shown since the observation period is unknown. Solid regions denote intervals where recorded data are available. Dashed red lines indicate seizure onsets (UEO) as marked by an epileptologist.

The data were expertly and independently marked to indicate EEC and UEO times, and 15-minute epochs were extracted for each patient. Baseline epochs were selected randomly from the entire patient record without regard to patient state-of-consciousness. For this experiment, a baseline epoch was defined as a 15-minute continuous recording period for which all of the known seizure onsets occurred at least

three hours before, or after, the baseline interval. A practical consequence of this baseline definition is that the initial- and final 1.5 hours of a continuous EEG recording period are excluded from consideration as baseline epochs (i.e. it is unknown if seizure onsets occurred at the boundaries of the recording period). An ictal epoch was selected for each temporal lobe seizure that a patient exhibited. Patients 7 and 10 exhibited some seizures with extra-temporal focal regions: those seizures were excluded from further analysis. Ictal epochs were extracted in a consistent manner such that the UEO occurred at a 10-minute offset within the epoch, allowing for analysis of both pre-ictal and post-ictal regimes. All baseline and ictal epochs were expertly reviewed to confirm the absence of recording artifacts. Recording artifacts and extra-temporal seizure foci resulted in exclusion of 18 seizures. The final dataset, summarized in Table 4-1, consisted of 29 seizure epochs and 49 baseline epochs.

Table 4-1. Summary of dataset selected for analysis. Each epoch (baseline or seizure) contains a 15-minute continuous IEEG recording. Note that selection criteria result in exclusion of 18 seizures (38 % of the subset). The focus channel listed represents the channel declared as the focus most frequently by an epileptologist.

Patient	Baseline Epochs	Seizure Epochs	Seizures	Focus Channel
1	14	5	6	LT2
4	10	6	9	RT2
5	6	5	5	RT2
7	9	7	13	LT2
10	10	6	14	LT3
All	*49*	*29*	*47*	-

It should be noted that the distinction between the normal (i.e. baseline) and novel (i.e. seizure) classes made in this research results in a *class imbalance problem*. A class imbalance problem exists when the relative fraction of instances from one class (e.g. baseline epochs) far exceeds the relative fraction of instances from another class (e.g. seizure epochs). While the dataset described in this section is not indicative of a class imbalance problem (the relative fractions are 0.37 and 0.63 for seizure- and baseline epochs, respectively), a consideration of the entire corpora does reveal a class imbalance problem. Approximately 1050 hours of data were reviewed and 106 seizure onsets noted: assuming a typical ictal/post-ictal event duration of 30 minutes, the estimated fraction of data associated with the novelty class is less than 0.05. Class imbalance problems generally require special consideration for classifier training (e.g. class weighting, or specialized sampling) to reduce classifier bias. In this work, the class imbalance issue is addressed through the specification of the ν parameter of the support vector machine employed (c.f. 3.4.1, 3.4.2). This parameter encodes an *a priori* estimate of the degree of novelty anticipated.

One final aspect of the dataset merits emphasis: the data are marked subjectively by two independent, fellowship trained and boarded epileptologist. This directly affects the performance assessment of any algorithm with regard to detection latency estimates and FPR estimates. The former are particularly sensitive to accurate, consistent marking, which may not be possible in non-focus channels. The latter are affected most significantly by seizures (especially subclinical seizures) that are not marked by an epileptologist. For instance, D'Alessandro [4] reported that many baseline segments

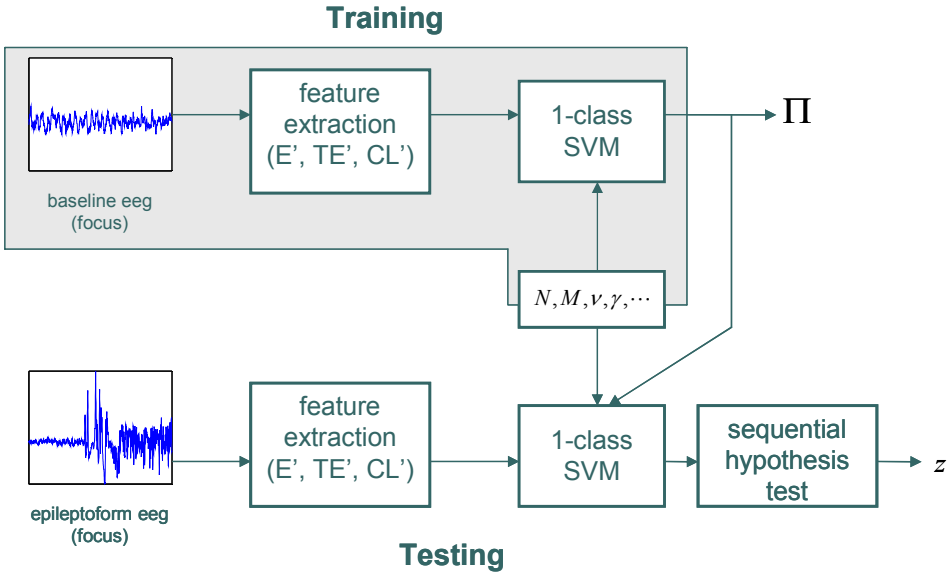

Fig. 4-2. An architecture for seizure analysis using a novelty detection framework. A small, randomly selected fraction of 15-minute baseline epochs were used for training. Ictal epochs were subsequently analyzed and the performance of the framework was evaluated.

flagged by a seizure detection/prediction algorithm were marked by an epileptologist, upon subsequent review, as exhibiting significant clinical or subclinical activity

4.1.1.2 Training

A novelty detection classifier was trained on the baseline data as shown in Figure 4-2. Each baseline segment was block-processed in 1-second frames with 0.5-second frame overlap. These block-processing window- and overlap durations were selected based upon previous researchers' works [2-4, 14]. Feature vectors for each frame (cf. 3.3.1.1) were extracted from baseline epochs. A total of 10,800 feature vectors were

presented to the 1-class SVM classifier for each baseline epoch. Additional system parameters were specified during training: the Gaussian radial basis function ($\gamma = 1.0$) was selected as the kernel function, the outlier fraction ($\upsilon = 0.1$) was chosen to be consistent with the estimated class fraction for the novelty class, the novelty detection threshold was set ($p = 0.8$) to produce a low false alarm rate, and the number of frames used to estimate p was fixed ($N = 20$). The resulting classifier model, Π, was stored for subsequent use in testing.

4.1.1.3 Testing

Testing was conducted using ictal segments containing seizures as shown in Figure 4-2. Feature vectors were extracted from ictal segments in a manner identical to the training case. A sequential hypothesis test was performed on the classifier output to provide a frame-wise continuous estimate of the probability of the detector being in the novelty state. The sequential hypothesis test, in conjunction with the novelty detection threshold, provides a mechanism for controlling the false alarm rate. A novelty event was declared whenever the indicator variable, z, exceeded the threshold, C.

4.1.1.4 Performance Assessment

Testing was performed using LOO-CV (leave-one-out-cross-validation, c.f. 3.5.1) to estimate four key performance criteria: sensitivity, mean detection latency, early-detection fraction, and FPR. The FPR was not deemed to be a critical performance metric for system evaluation. There are four justifications for this decision: (1) the present technique is an unsupervised learning technique that detects novelties; it is known that novelties are not restricted to ictal events; (2) it is hypothesized that the FPR can be

reduced by extending the detection system, possibly by incorporating a traditional classifier to hierarchically discriminate novelty frames; (3) recent trends in building a seizure prevention device have led to growing acceptance of a hyperdetection philosophy among epilepsy researchers; (4) the present technique is useful without regard to FPR as a new, offline, post-processing tool for epilepsy researchers.

4.1.2 Results

The system was evaluated using focus channel data and cross-validation (c.f. 3.5.1) to estimate the performance parameters.

4.1.2.1 Detector Performance – Sensitivity

The novelty detection system detected 100% of all onsets with no false negatives over the range of detector persistence settings. This corresponds to a sensitivity of 100%.

Figure 4-3 shows the results of processing a single baseline epoch for Patient 5. As expected, no portion of this epoch is declared as a novelty. The low-amplitude, high-frequency (or white noise) region near the end of the baseline epoch biases the detector state, but does not exceed the novelty threshold. This low-amplitude, high-frequency behavior is typical of early electrographic changes evident in many seizure onsets. Figure 4-4 shows the results of processing a single ictal epoch for Patient 5. As expected, the seizure onset is detected. The sharp rise in detector state is typical of seizure detections. It is interesting to observe that the novelty detector declares the post-ictal period as novel, also.

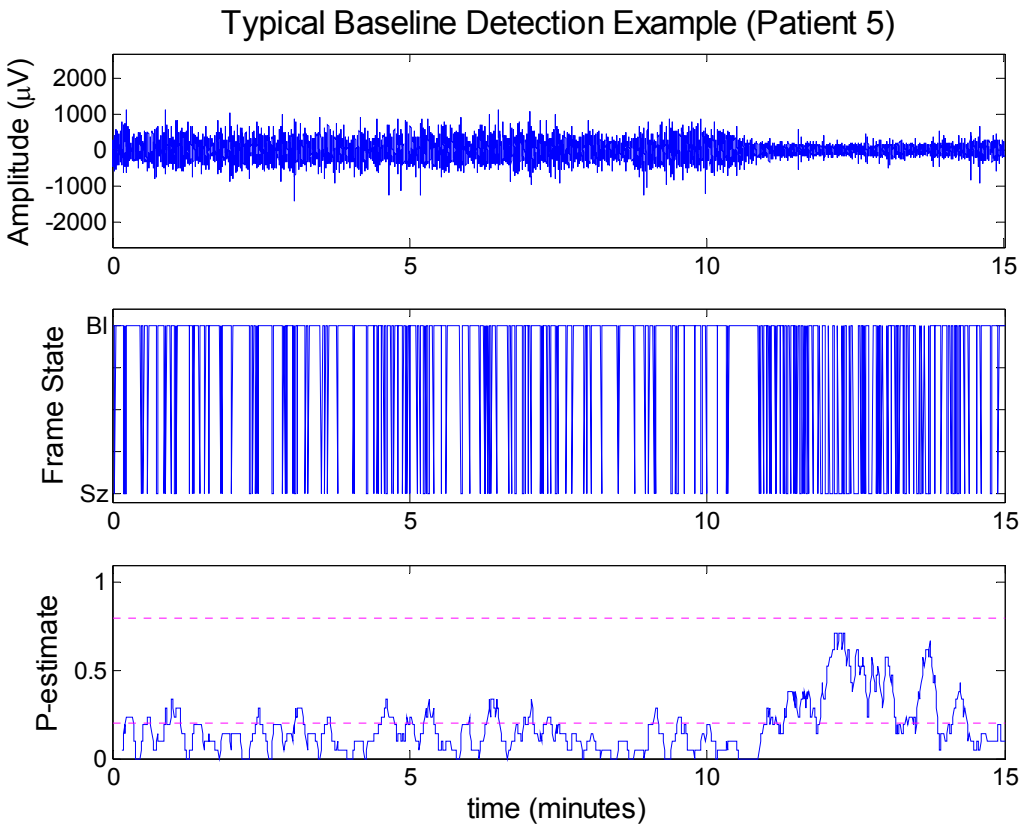

Fig. 4-3. A typical baseline epoch. (*Top*) IEEG signal, (*Middle*) frame-wise output of the novelty detector, z, (*Bottom*) probability estimate of being in the novelty state (probability thresholds at 0.8 and 0.2 are shown for convenience).

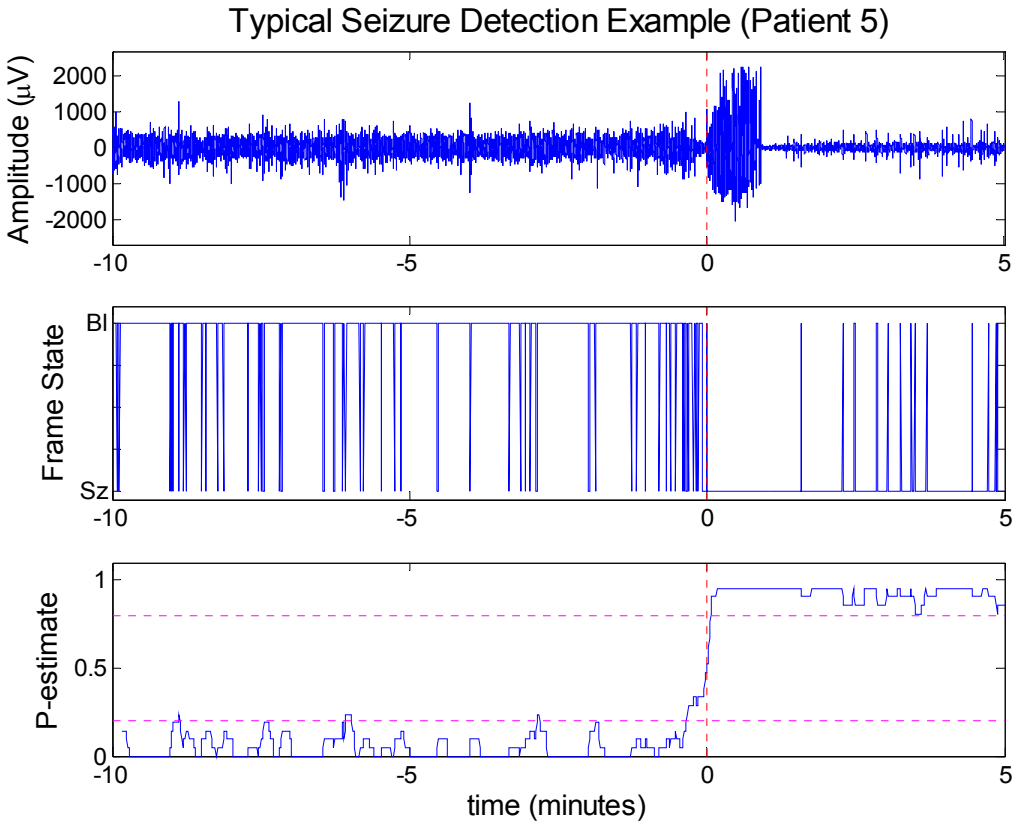

Fig. 4-4. A typical ictal epoch. (*Top*) IEEG signal. The EEC is visible as the beginning of the pinched region prior to the high-amplitude seizure onset. The UEO occurs at time zero, (*Middle*) frame-wise output of the novelty detector, z, (*Bottom*) probability estimate of being in the novelty state (probability thresholds at 0.8 and 0.2 are shown for convenience). The detector has a latency of about 3 seconds in this example.

4.1.2.2 Detector Performance – Detection Latency

The effect of detector latency was characterized for six persistence settings. Figure 4-5 summarizes the results. A sharp decline in the mean detection latency is observed between persistence values of T = 60 seconds and T = 120 seconds. This

decline continues in an almost asymptotic fashion. The latency-persistence behavior is predominantly due to a subset of seizure events that seem to be early-detectable.

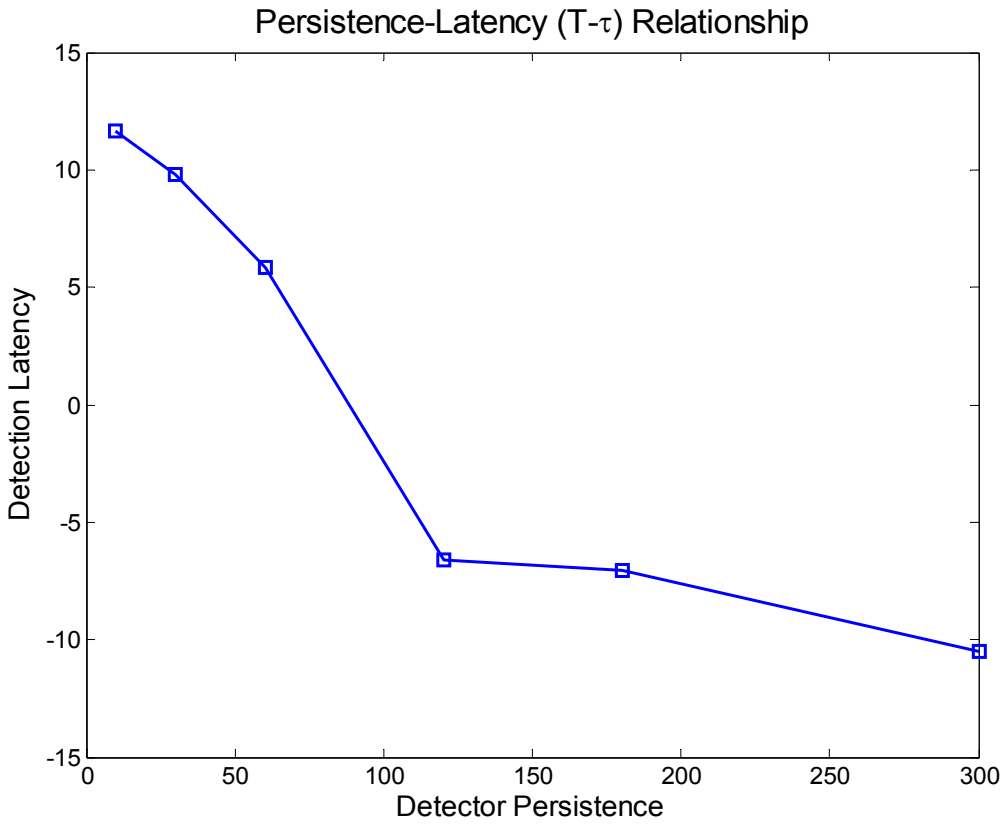

Fig. 4-5. Relationship between persistence and mean detection latency.

Detailed latency results for a range of persistence are presented in Table 4-2.

Table 4-2. Mean detection latency results (in seconds) for the static novelty detector computed using LOO-CV. System parameters, Π, were fixed while the detector refractory period was varied. Negative latency values (*highlighted*) indicate an early detection.

Patient	Sz	Detector Persistence, T (seconds)					
		10	*30*	*60*	*120*	*180*	*300*
1	1	10.75	10.75	**-50.83**	**-50.83**	**-50.83**	**-50.83**
	2	33.83	**-29.25**	**-29.25**	**-77.08**	**-77.08**	**-77.08**
	3	**-2.17**	**-2.17**	**- 2.17**	**-68**	**-68**	**-68**
	4	8	8	8	**-91.17**	**-91.17**	**-91.17**
	5	5.17	5.17	5.17	**-90.42**	**-90.42**	**-90.42**
4	4	6.85	6.85	6.85	6.85	6.85	6.85
	5	6.95	6.95	6.95	6.95	6.95	6.95
	6	7.05	7.05	7.05	7.05	7.05	7.05
	7	4.75	4.75	4.75	4.75	4.75	4.75
	8	7.85	7.85	7.85	7.85	7.85	7.85
	9	7.8	7.8	7.8	**-0.45**	**-0.45**	**-0.45**
5	1	4.75	4.75	4.75	4.75	4.75	4.75
	2	5.75	5.75	5.75	5.75	5.75	5.75
	3	3.333	3.333	3.333	3.333	3.333	3.333
	4	3.083	3.083	3.083	3.083	3.083	3.083
	5	**-4**	**-4**	**-4**	**-4**	**-4**	**-4**
7	1	156.1	156.1	156.1	156.1	156.1	156.1
	2	2.556	2.556	**-14.67**	**-14.67**	**-27.28**	**-39.56**
	3	6	3.556	2.056	**-110.4**	**-110.4**	**-188.6**
	4	3.444	3.444	3.444	3.444	3.444	3.444
	7	4.333	4.333	4.333	4.333	4.333	4.333
	8	3.5	3.5	3.5	3.5	3.5	3.5
	9	6.333	6.333	6.333	6.333	6.333	6.333
10	2	**-7.444**	**-11.22**	**-11.22**	**-11.22**	**-11.22**	**-11.22**
	3	8.444	8.444	8.444	8.444	8.444	8.444
	4	8.611	8.611	8.611	8.611	8.611	8.611
	6	9.944	9.944	**-35.28**	**-35.28**	**-35.28**	**-35.28**
	7	5.389	5.389	5.389	5.389	5.389	5.389
	9b	**-7.444**	**-11.22**	**-11.22**	**-11.22**	**-11.22**	**-11.22**
(all)		11.65	9.82	5.83	**-6.60**	**-7.08**	**-10.51**

Results from Table 4-2 suggest that changes in detection latency due to increases in detector persistence are seizure-specific. It is hypothesized that two subclasses of seizures are actually present in the dataset: those that are merely detectable, and those

that may be predictable. These classes of seizures appear to be patient-dependent, e.g. Patients 4 and 5 seizures are not predictable. Otherwise, the detection results appear to be quite stable and independent of that particular baseline epoch used for training. This confirms that feature vectors derived from ictal epochs are significantly different from those associated with normal brain rhythms. The estimated mean detection latency is -7.08 seconds for a persistence value of T = 180 seconds.

4.1.2.3 Detector Performance -- Early Detection Fraction

As noted in 4.1.2.2, an estimated 41% of seizures are predictable for a persistence of T = 180 seconds. Figure 4-6 shows a typical early detection of -40 seconds.

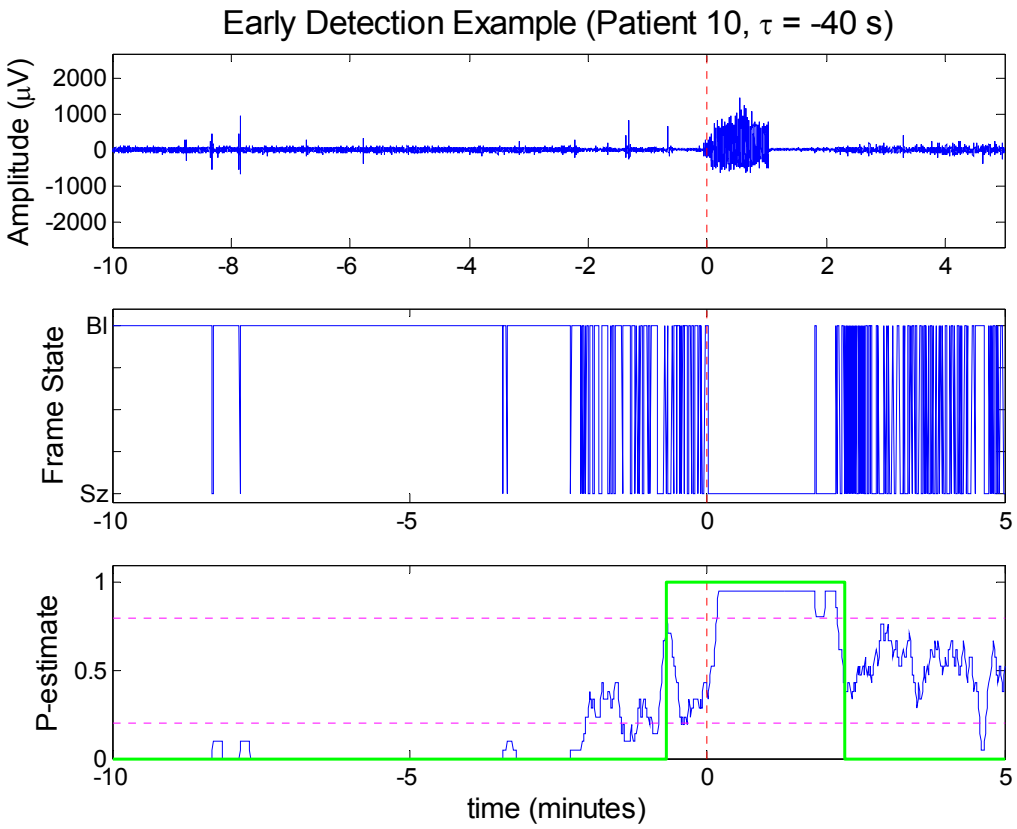

Fig. 4-6. An example of an early-detection. (*Top*) IEEG signal, (*Middle*) frame-wise output of the novelty detector, z, (*Bottom*) probability estimate of being in the novelty state. The green indicator function represents the output of the novelty detection system with persistence.

Note the detector state "builds up" approaching the UEO. Figure 4-7 shows an expanded section of Figure 4-6. The detector state appears to be sensitive to decreasing amplitude and increasing frequency changes in the IEEG. This is consistent with epileptologist use of visual cues for marking chirps.

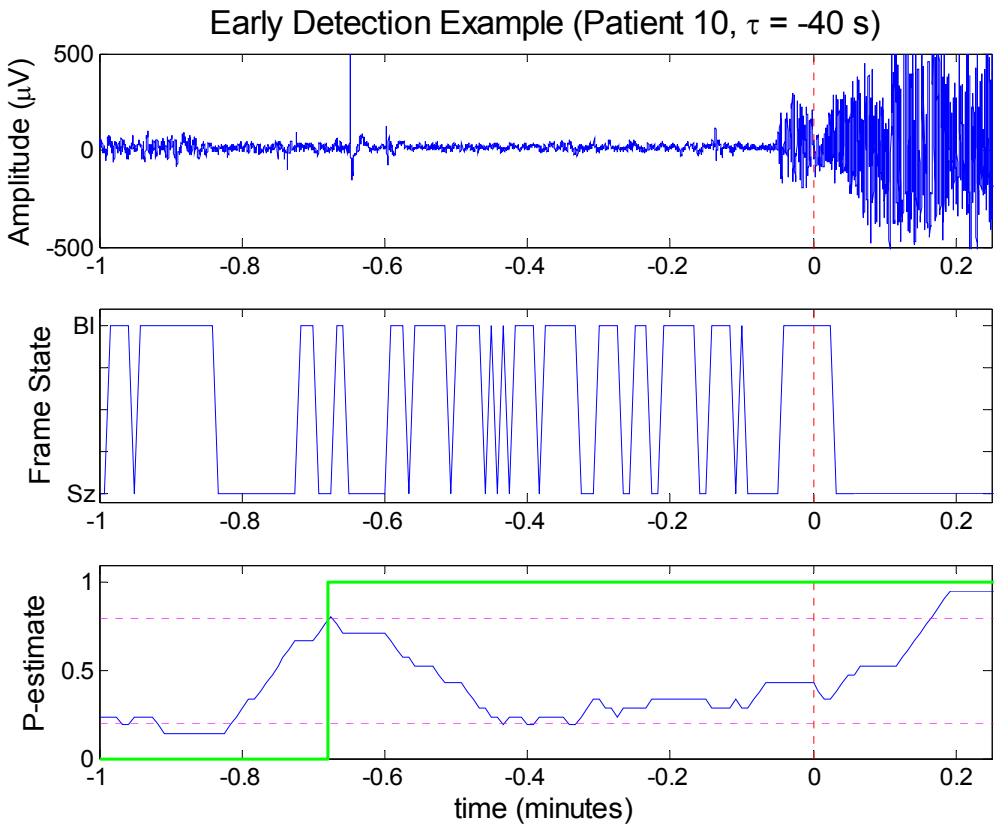

Fig. 4-7. An example of an early detection event (from Figure 4-6).

An example of a possible data integrity issue manifested as an early detection is shown in Figure 4-8. The seizure begins diffusely, and it is not clear that this is a stereotypical focus-channel seizure. The UEO indicated by the epileptologist appears late, resulting in an apparent detection latency of -59.5 seconds.

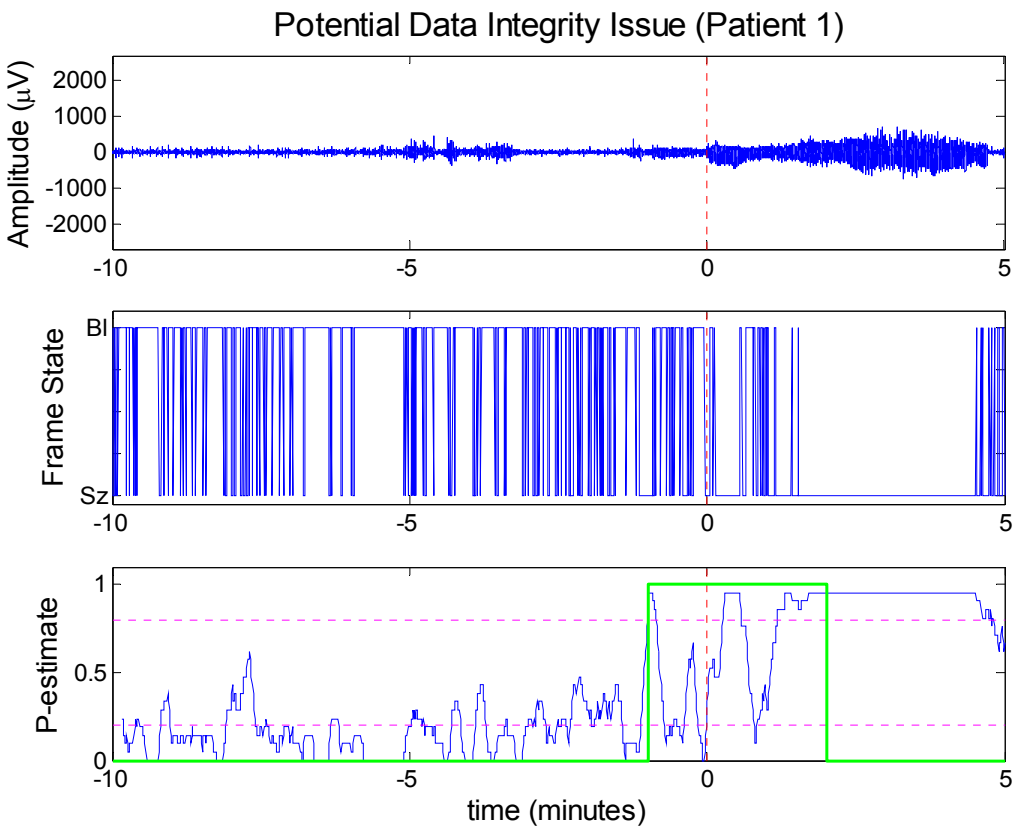

Fig. 4-8. An example illustrating possible data integrity issues.

4.1.2.4 Detector Performance – FPR

The detector FPR performance is summarized in Table 4-3 and Figure 4-9.

Table 4-3. Patient-specific effects of detector persistence on FPR. Each persistence value, T, contains two columns of numbers: (*left*) number of false positives on ictal epochs, (*right*) number of false positives on baseline epochs.

Patient	Detector Persistence, T (in seconds)											
	10 s		*30s*		*60s*		*120s*		*180s*		*300s*	
1	0	38	0	29	0	24	0	0	0	0	0	0
4	33	4	23	4	19	4	14	2	10	2	9	2
5	0	0	0	0	0	0	0	0	0	0	0	0
7	26	14	15	14	15	11	6	11	0	10	0	8
10	15	14	8	11	0	9	0	6	0	5	0	4
All	**83**	**61**	**56**	**48**	**43**	**39**	**8**	**31**	**2**	**25**	**2**	**19**

.

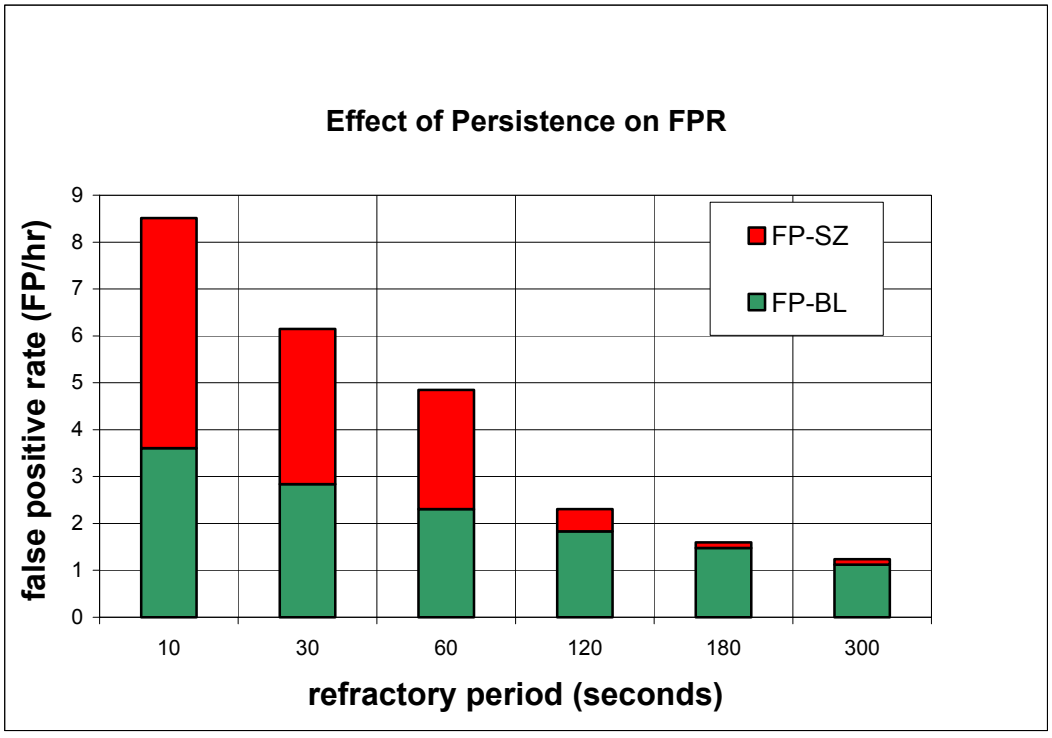

Fig. 4-9. False positive rates for ictal (*red*) and baseline (*green*) epochs estimated from static experiments using cross-validation.

The FPR is approximately exponentially distributed and consists of two components: false positives that occur on ictal epochs, and false positive that occur on baseline epochs. The effect of increasing persistence is a dramatic reduction in the false positive rate over ictal epochs. However, since the ictal epochs represent only 5% of the total data, this improvement in FPR quickly becomes negligibly small compared to the FPR contribution from baseline epochs. Estimated FPRs account for the class imbalance between ictal and baseline data. The asymptotic false positive rate is estimated to be approximately 1.25 Fph. Figure 4-10 shows a typical false positive for a baseline epoch.

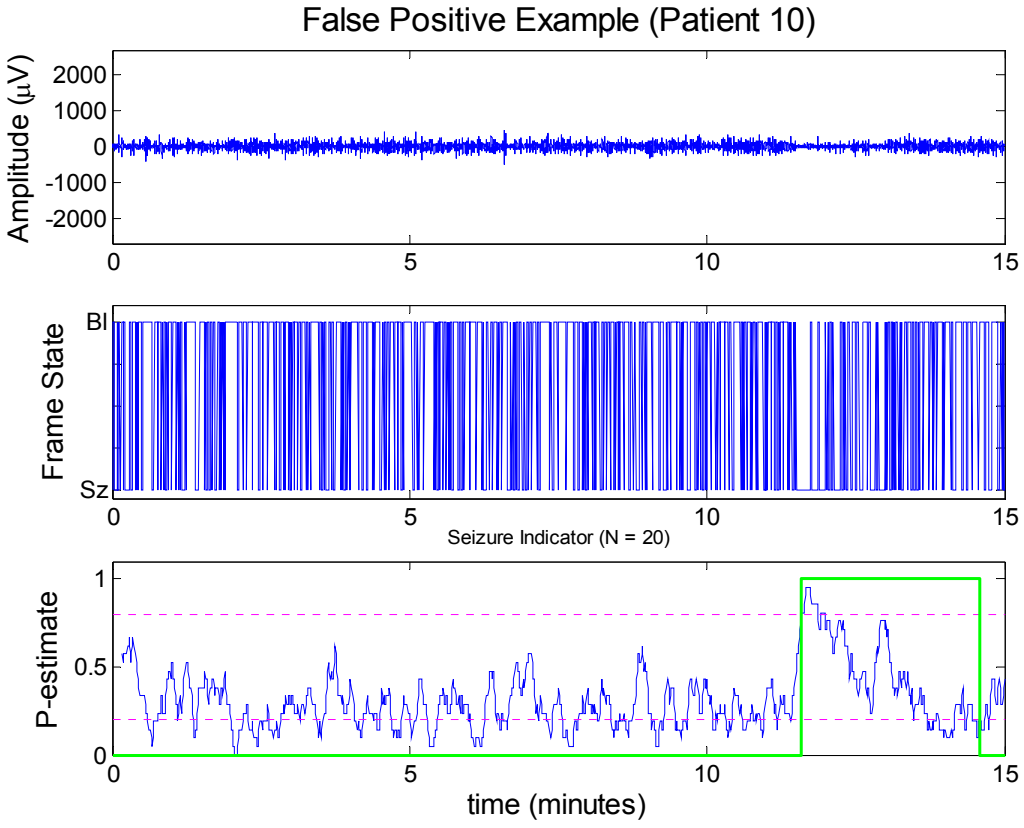

Fig. 4-10. An example of a typical baseline-occurring false positive event.

Figure 4-11 shows the baseline epoch that exhibited the worst-case false positive behavior. Note that this baseline epoch appears atypical and exhibits large-amplitude spiking. The presence of numerous novelty events perpetuates a near-continual firing of the novelty detector even with persistence.

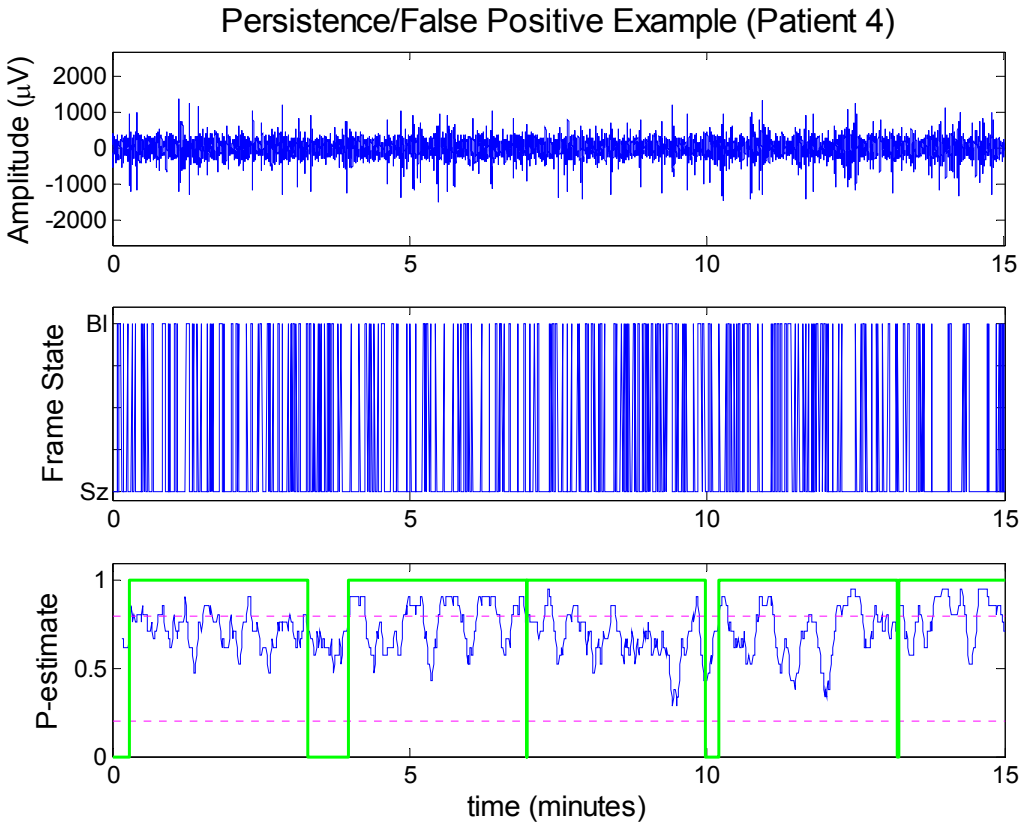

Fig. 4-11. The worst-case example of false positives considered in the data set.

4.1.3 Discussion

A clear design tradeoff exists when specifying the detector persistence for the system. A larger value of persistence decreases the FPR, but increases the mean detection latency. No specific objective function was formulated to determine an optimal persistence setting, but the data suggests that a reasonable persistence setting lies between T = 120 seconds and T = 180 seconds. For a persistence setting of T = 180 seconds, the static detector achieved 100% sensitivity, -7.08 mean detection latency, and 41% early-detection fraction at 1.57 Fph. However, the population size for estimation of false negative rate—a key contributor to the overall sensitivity rating—was small. Still, the results obtained are significant and justify further investigation into the behavior of the novelty detector for different parameter values, and the development of an online detector.

Direct comparison with competing methods is difficult due to the different datasets analyzed, features employed, and algorithmic implementations. Echauz et al. [31] optimized a seizure detector using data from the same corpora. Their detector relied on three similar features, but was heuristically tuned in a patient-specific manner over more than 1265 hours of recordings. Their detector is the basis of a commercial, implantable device detector, and represents the state-of-the-art in seizure detection. Their results were: 98.4% sensitivity, mean detection latency of 5 seconds, and 1.59% early-detection fraction at 0.013 Fph. The results of the novelty detection system compare favorably, yielding a higher sensitivity, earlier mean detection latency, and much greater early-detection fraction. Only FPR is worse; however, this may be acceptable for a

seizure prevention device (c.f. 4.1.1.5), or may represent previously unidentified novelty events.

A practical advantage of the proposed method over previous classification approaches is the formulation of the detection threshold. The theoretical and empirical properties of the 1-class SVM, combined with a simple maximum likelihood estimator, provide a straightforward method for controlling the tradeoff between detection latency and false alarm rate. The introduction of detector persistence is an intuitive and straightforward method for improving the FPR and characterizing the detector over a range of detection horizons. The present method is computationally efficient due to the use of the SVM. Only 1100 feature vectors are required for analysis of a single 15-minute epoch. The novelty detector requires only a single epoch of baseline data for training per patient, and it does not require patient-specific threshold tuning or feature selection. The zero FNR of the novelty detector was surprising since the training data represented varying (and unknown) patient sleep states; evaluation on larger data sets will provide a better estimate of the true FNR.

4.2 Parameter Tuning

The system as presented has five major parameters of interest: $\Pi = (\nu, \gamma, p, N, T)$. Results from 4.1.2 demonstrated that the novelty detector robustly detects seizures for the default model parameters $\Pi = (0.1, 1.0, 0.8, 20, 180)$, and characterized the effects of varying the persistence, T. A series of additional static detector experiments were performed to characterize the effects of other parameter variations on system performance. In each experiment, a single model parameter was selected and varied over

a range of values with the remaining parameters fixed at default values. Performance metrics were estimated using LOO-CV. Note that in the present section FNR is considered separately instead of sensitivity since FNR was determined to be solely responsible for detector sensitivity performance. Also note that these experiments were not aimed at optimizing detector performance. Such a task is very challenging because of the difficulty in formulating a universal objective function for optimization. A pseudocode representation of the experimental procedure is shown in Figure 4-12.

$\Pi = (\nu, \gamma, p, N, T) = \bar{\boldsymbol{\pi}}$
T = 180 seconds
For each parameter, π_i, in Π
  For each patient, p_k, in dataset
    For each baseline epoch, E_m^{BL}
      Train novelty detector for new $\pi_i = c$
        For each seizure epoch, E_m^{SZ}
          Estimate performance metrics
End

Fig. 4-12. Pseudocode overview of the parameter tuning experiments.

4.3.1 Effects on FPR

4.3.1.1 The Effect of ν

The results of varying ν, the novelty parameter, on the estimated FPR are presented in Table 4-4. The FPR clearly increases for all patients as ν increases, especially for small values of the parameter. The estimated FPR values show patient

variability: Patient 5 has an especially low FPR. The FPR values presented in Table 4-4 are higher than the results from 4.1.2. This is due to different estimation procedures: no correction was made here for class imbalance, so the false positives occurring on ictal segments dominate the FPR estimate and yield higher values than those reported in 4.1.2.

Table 4-4. The effect of varying ν on the estimated FPR.

ν	FPR_1	FPR_4	FPR_5	FPR_7	FPR_{10}
0.025	2.19	2.98	0.29	0.95	5.43
0.050	3.03	5.30	0.73	1.95	6.16
0.075	3.65	6.83	1.21	2.49	7.18
0.100	4.28	7.24	1.95	3.27	7.14
0.125	5.13	7.31	2.88	3.56	7.25
0.150	5.91	7.37	3.59	4.26	7.44
0.175	5.82	7.48	4.02	4.85	7.32
0.200	5.93	7.47	4.25	5.05	7.62
0.250	7.18	7.85	4.49	5.98	7.78

4.3.1.2 The Effect of γ

The results of varying γ, the similarity parameter, on the estimated FPR are presented in Table 4-5. The FPR is relatively insensitive to changes in γ, but for some patients a value of 1.0 or 3.0 produces the lowest dataset FPR. The estimated FPR values show some patient variability, especially for Patients 5 and 7.

Table 4-5. The effect of varying γ on the estimated FPR.

γ	FPR_1	FPR_4	FPR_5	FPR_7	FPR_{10}
0.10	5.56	5.59	4.4	3.56	6.17
0.25	5.81	6.28	3.72	3.42	6.88
0.50	5.21	7.10	3.59	3.77	7.50
1.0	5.13	7.31	2.88	3.56	7.25
3.0	4.99	6.98	2.32	3.94	6.03
10.0	5.55	7.40	2.36	3.84	5.95

4.1.3.3 The Effect of p

The results of varying p, the novelty state probability threshold parameter, on the estimated FPR are presented in Table 4-6. The FPR is insensitive to changes in p, below a critical value of about 0.65. Above this critical value the FPR decreases sharply as the parameter increases.

Table 4-6. The effect of varying p on the estimated FPR.

p	FPR_1	FPR_4	FPR_5	FPR_7	FPR_{10}
0.5	8.56	6.95	5.94	8.37	7.46
0.65	6.71	7.52	4.43	5.41	7.45
0.80	5.13	7.31	2.88	3.56	7.25
0.85	4.05	7.40	2.11	3.24	7.26
0.9	2.79	6.62	1.49	2.30	6.55

4.1.3.4 The Effect of N

The results of varying N, the number of frames for estimating p, on the estimated FPR are presented in Table 4-7. As expected, the relationship between N and FPR is strongly linear (with negative slope): the more frames considered novel, the greater the lower the probability of generating a false positive.

Table 4-7. The effect of varying N on the estimated FPR.

N	FPR_1	FPR_4	FPR_5	FPR_7	FPR_{10}
15	6.17	8.34	3.47	4.36	8.72
20	5.13	7.31	2.88	3.56	7.25
30	3.99	6.58	1.94	2.87	6.12
50	2.23	5.16	1.03	1.83	4.29
100	0.95	3.56	0.71	1.01	3.36

4.1.3.5 Summary of Parameter Effects

A summary of the effects of model parameter changes on FPR is graphically depicted in Figure 4-13. The monotonic behavior of ν-, p-, and N-induced changes is expected. The detector is relatively insensitive to γ.

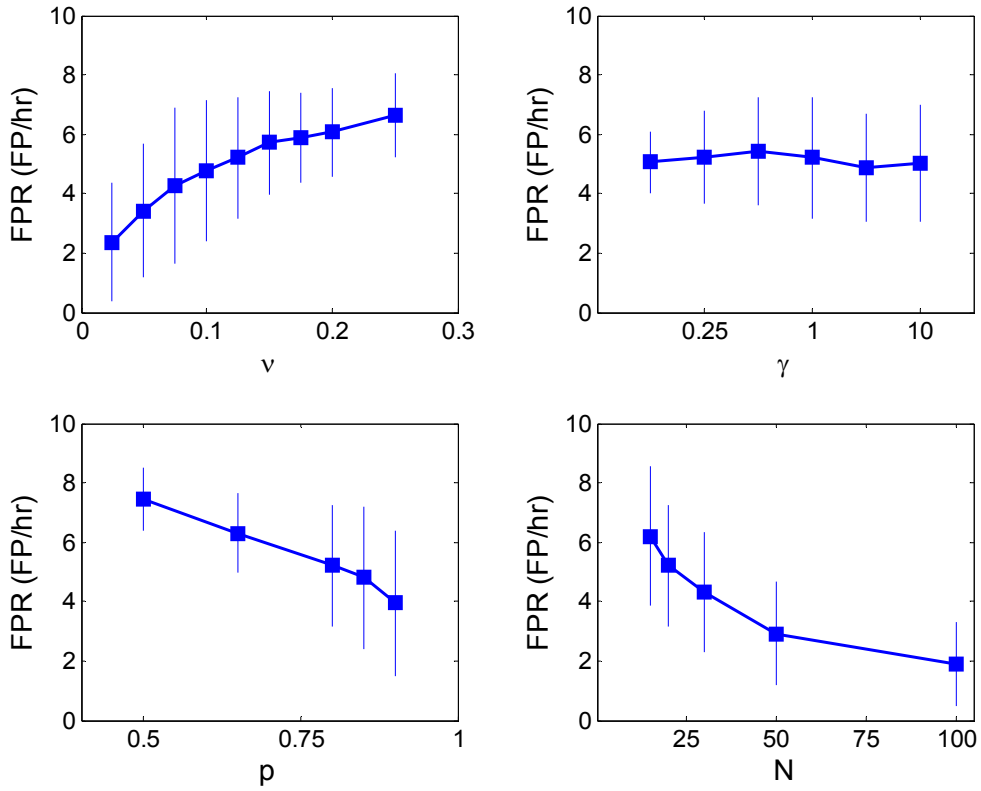

Fig. 4-13. Summary of parameter effects on estimated FPR of the novelty detector. Plots show mean FPR values and 1-std error bars. Note that the abscissa for gamma is nonlinearly scaled for display purposes.

The following all lower the overall FPR of the system: increasing p or N, or decreasing ν.

4.3.2 Effects on FNR

Since the default static novelty detector correctly detects all onsets, the FNR is the dominant term in the sensitivity calculation. The effects of parameter changes on FNR are presented in this section.

4.3.2.1 The Effect of ν

The results of varying ν, the novelty parameter, on the estimated FNR are presented in Table 4-8. Patient 5 uniquely exhibits no false negatives over the range of ν; the other patients generally show a FPR that is proportional to the parameter, but approaching a patient-specific asymptote. The FNR is non-negative for all cases. The FNR is non-negative for the default static detector parameter values because of the increased number of baseline segments used for cross-validation.

Table 4-8. The effect of varying ν on the estimated FNR.

ν	FNR_1	FNR_4	FNR_5	FNR_7	FNR_{10}
0.025	0	0	0	0.0317	0.2833
0.050	0.0667	0.0500	0	0.0794	0.3833
0.075	0.1000	0.0667	0	0.1905	0.3167
0.100	0.1000	0.1000	0	0.2540	0.3000
0.125	0.1000	0.1500	0	0.3333	0.3167
0.150	0.0667	0.1500	0	0.3968	0.3667
0.175	0.0667	0.1500	0	0.4286	0.3667
0.200	0.1000	0.1833	0	0.4286	0.4500
0.250	0.1667	0.2333	0	0.4444	0.4833

4.3.2.2 The Effect of γ

The results of varying γ, the similarity parameter, on the estimated FNR are presented in Table 4-9. Again, Patient 5, exhibits no false negatives. There is a weak correspondence between increasing γ and increasing FNR.

Table 4-9. The effect of varying γ on the estimated FNR.

γ	FNR_1	FNR_4	FNR_5	FNR_7	FNR_{10}
0.10	0	0.1167	0.0333	0.2222	0.1833
0.25	0.0333	0.0833	0	0.2222	0.2167
0.50	0.0667	0.1167	0	0.3333	0.3167
1.0	0.1000	0.1500	0	0.3333	0.3167
3.0	0.0667	0.1500	0	0.3333	0.4333
10.0	0.1333	0.1833	0.0333	0.3810	0.6000

4.3.2.3 The Effect of p

The results of varying p, the novelty state probability threshold parameter, on the estimated FNR are presented in Table 4-10. Patient 5 exhibits no false positives over the range of the parameter. The FNR is insensitive to changes in p, below a critical value of about 0.65. Above this critical value the FNR decreases sharply as the parameter increases.

Table 4-10. The effect of varying p on the estimated FNR.

p	FNR_1	FNR_4	FNR_5	FNR_7	FNR_{10}
0.5	0.2333	0.3167	0	0.4762	0.6667
0.65	0.2000	0.2167	0	0.3968	0.4833
0.80	0.1000	0.1500	0	0.3333	0.3167
0.85	0.1000	0.1000	0	0.2063	0.3167
0.9	0.0667	0.0333	0	0.0794	0.2500

4.3.2.4 *The Effect of N*

The results of varying N, the number of frames for estimating p, on the estimated FNR are presented in Table 4-11. This FNR appears relatively insensitive to this parameter.

Table 4-11. The effect of varying N on the estimated FNR.

N	FNR_1	FNR_4	FNR_5	FNR_7	FNR_{10}
15	0.1	0.15	0	0.3492	0.3167
20	0.1	0.15	0	0.3333	0.3167
30	0	0.1333	0.0333	0.254	0.3667
50	0.0333	0.15	0	0.3175	0.3
100	0.0333	0.1167	0	0.254	0.2833

4.3.2.5 *Summary of Parameter Effects*

A summary of the effects of model parameter changes on FNR is graphically depicted in Figure 4-14. The monotonic behavior of ν-, p-, and γ-induced changes is expected. The detector is relatively insensitive to N.

.

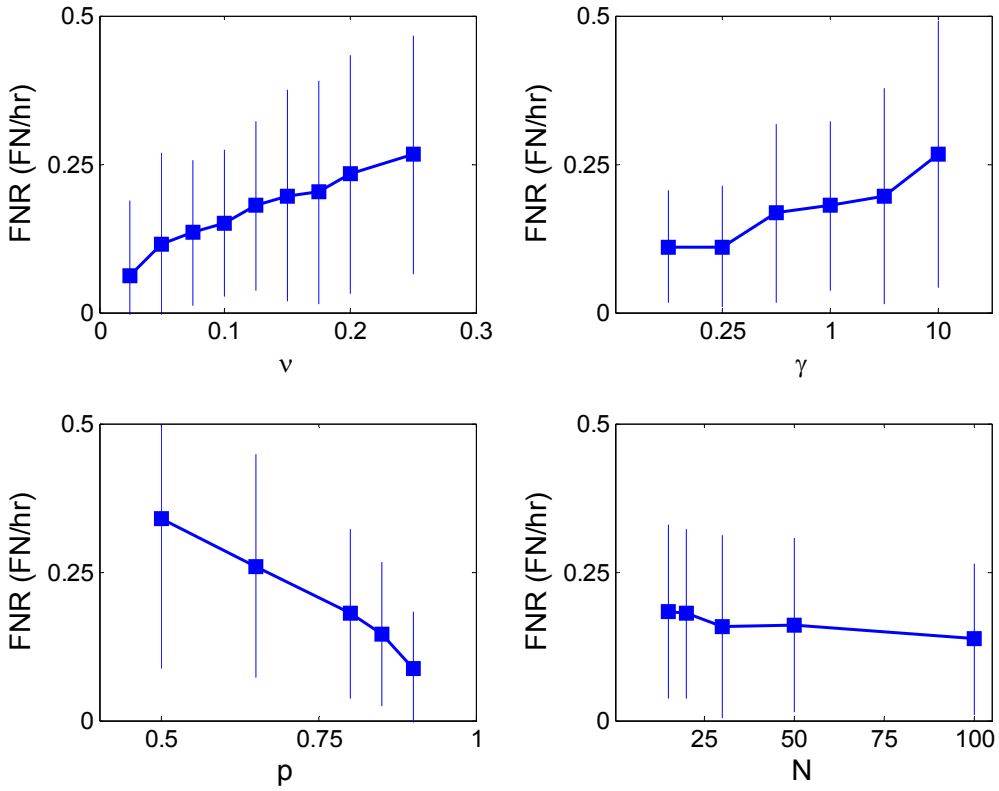

Fig. 4-14. Summary of parameter effects on estimated FNR of the novelty detector. Plots show mean FNR values and 1-std error bars. Note that the abscissa for gamma is nonlinearly scaled for display purposes.

The following all lower the overall FNR of the system: increasing p, or decreasing γ or ν.

4.3.3 Effects on Detection Latency

The effects of parameter changes on mean detection latency are presented in this section.

4.3.3.1 The Effect of ν

The results of varying ν, the novelty parameter, on the estimated mean detection latency are presented in Table 4-12. A general trend is observable: the higher the value of ν, the earlier the detection. This agrees with our understanding of the novelty detector. By increasing ν, a greater fraction of the data is accepted as novel. Anomalous results—perhaps local extrema—occur for some patients (e.g. $\nu = 0.1$ for Patient 4). Patients 1, 4, and 10 exhibit this behavior. This may be related to the mean seizure rate of patients.

Table 4-12. The effect of varying ν on the estimated mean detection latency (in seconds).

ν	τ_1	τ_4	τ_5	τ_7	τ_{10}
0.025	-68.0	7.38	2.75	3.75	-19.3
0.050	-78.0	6.13	2.50	-1.50	5.5
0.075	-90.5	3.50	1.25	-2.50	3.38
0.100	-94.0	3.38	0.00	3.25	-19.5
0.125	-78.3	5.25	-0.75	-20.0	-9.25
0.150	-70.5	1.50	-2.00	-23.5	-11.8
0.175	-70.3	3.00	-3.00	-34.0	-13.1
0.200	-70.5	-8.38	-3.00	-48.0	-5.38
0.250	-122.5	-9.63	-3.50	-75.3	-27.8

4.3.3.2 *The Effect of γ*

The results of varying γ, the similarity parameter, on the estimated mean detection latency are presented in Table 4-13. A local minimum in latency is observed for three patients for γ between 1.0 and 3.0.

Table 4-13. The effect of varying γ on the estimated mean detection latency (in seconds).

γ	τ_1	τ_4	τ_5	τ_7	τ_{10}
0.10	-43.5	6	-3	-3	-8.5
0.25	-91.0	5.75	-3	-31	-2.38
0.50	-89.5	5.5	-2	-19.5	3.25
1.0	-78.3	5.25	-.75	-20	-9.25
3.0	-94.8	4.75	-.75	-24	-20.88
10.0	-95.3	5.25	2.25	-18	-39.25

4.3.3.3 *The Effect of p*

The results of varying p, the novelty state probability threshold parameter, on the estimated FNR are presented in Table 4-14. Lower parameter values generally correspond to earlier detections, but several exceptions are noted.

Table 4-14. The effect of varying p on the estimated mean detection latency (in seconds).

p	τ_1	τ_4	τ_5	τ_7	τ_{10}
0.5	-136.5	-40.6	-6.25	-4	-4.25
0.65	-71.5	-14.8	-4.75	-47.8	-35.8
0.80	-78.3	5.25	-0.75	-20	-9.25
0.85	-90.5	5	.75	-19.5	-59.5
0.9	-77.3	7.38	4.5	-30.5	-12.3

4.3.3.4 The Effect of N

The results of varying N, the number of frames for estimating p, on the estimated mean detection latency are presented in Table 4-15. Detection latency increases as the parameter is increased, as expected.

Table 4-15. The effect of varying N on the estimated mean detection latency (in seconds).

N	τ_1	τ_4	τ_5	τ_7	τ_{10}
15	-93	1	-3	-19	-29
20	-78.3	5.25	-0.75	-20	-9.25
30	-86.3	5.25	4.5	1.5	0.75
50	-66.5	14.5	11.75	-9.5	0.5
100	26	31.8	28.3	29	-11

4.3.3.5 *Summary of Parameter Effects*

A summary of the effects of model parameter changes on FNR is graphically depicted in Figure 4-15. The detection latency is modestly affected by changes in ν, p, and γ. The latency is strongly positively correlated N.

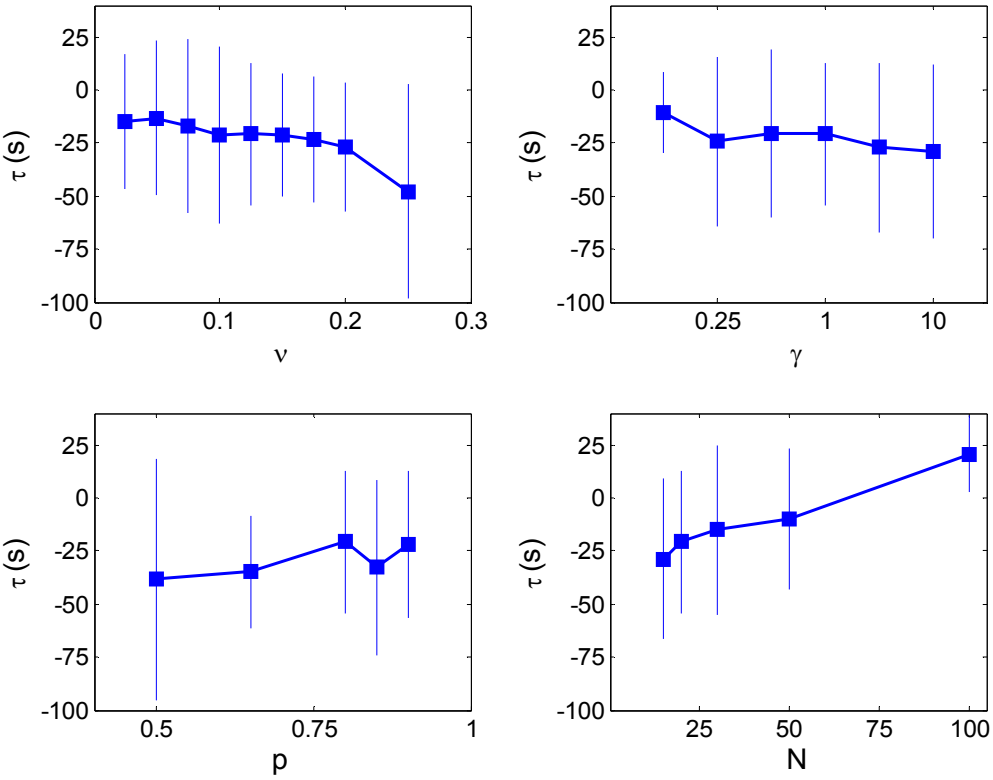

Fig. 4-15. Summary of parameter effects on estimated latency of the novelty detector. Plots show mean latency values and 1-std error bars. Note that the abscissa for gamma is nonlinearly scaled for display purposes.

4.3.4 Effects on Early Detections

4.3.4.1 The Effect of ν

The results of varying ν, the novelty parameter, on the estimated mean detection latency are presented in Table 4-16. The observable trend that a higher value of ν results in more early-detections is consistent with the function of the novelty detector. For a nominal value of $\nu = 0.125$ the early detection fraction is about 58.6%.

Table 4-16. The effect of varying ν on the number of early detections.

ν	ED_1	ED_4	ED_5	ED_7	FNR_{10}
0.025	5/5	0/6	1/5	2/7	3/6
0.050	5/5	0/6	1/5	4/7	1/6
0.075	5/5	2/6	1/5	4/7	2/6
0.100	5/5	2/6	2/5	3/7	4/6
0.125	5/5	1/6	3/5	5/7	3/6
0.150	5/5	1/6	4/5	5/7	5/6
0.175	5/5	1/6	4/5	5/7	5/6
0.200	5/5	4/6	4/5	7/7	4/6
0.250	5/5	3/6	5/5	7/7	5/6

4.3.4.2 The Effect of γ

The results of varying γ, the similarity parameter, on the estimated mean detection latency are presented in Table 4-17. Fewer early-detections for higher values of γ is consistent with the memorization effect of γ.

Table 4-17. The effect of varying γ on the number of early detections.

γ	ED_1	ED_4	ED_5	ED_7	ED_0
0.10	3/5	0/6	4/5	4/7	3/6
0.25	5/5	1/6	4/5	6/7	3/6
0.50	5/5	1/6	4/5	5/7	2/6
1.0	5/5	1/6	3/5	5/7	3/6
3.0	5/5	2/6	3/5	5/7	1/6
10.0	5/5	2/6	1/5	4/7	0/6

4.3.4.3 The Effect of p

The results of varying p, the novelty state probability threshold parameter, on the estimated FNR are presented in Table 4-18. The ability of the detector for early detections decreases sharply from $p = 0.65$ to $p = 0.8$.

Table 4-18. The effect of varying p on the number of early detections.

p	ED_1	ED_4	ED_5	ED_7	ED_{10}
0.5	5/5	5/6	5/5	7/7	3/6
0.65	5/5	5/6	5/5	7/7	5/6
0.80	5/5	1/6	3/5	5/7	3/6
0.85	5/5	1/6	2/5	5/7	4/6
0.9	5/5	0/6	1/5	6/7	4/6

4.3.4.4 *The Effect of N*

The results of varying N, the number of frames for estimating p, on the estimated mean detection latency are presented in Table 4-19. The ability of the detector for early detections decreases sharply from $N = 30$ to $N = 50$.

Table 4-19. The effect of varying N on the number of early detections.

N	ED_1	ED_4	ED_5	ED_7	ED_{10}
15	5/5	3/6	4/5	6/7	6/6
20	5/5	1/6	3/5	5/7	3/6
30	5/5	1/6	1/5	3/7	3/6
50	4/5	1/6	0/5	5/7	3/6
100	2/5	0/6	0/5	2/7	3/6

4.3.4.5 *Summary of Parameter Effects*

A summary of the effects of model parameter changes on mean detection latency is graphically depicted in Figure 4-16. The success at early detection is strongly dependent on ν, p, and N. Early detection fraction also appears to be patient-dependent. The following all increase the number of early detections generated by the system: increasing ν, or decreasing p or N.

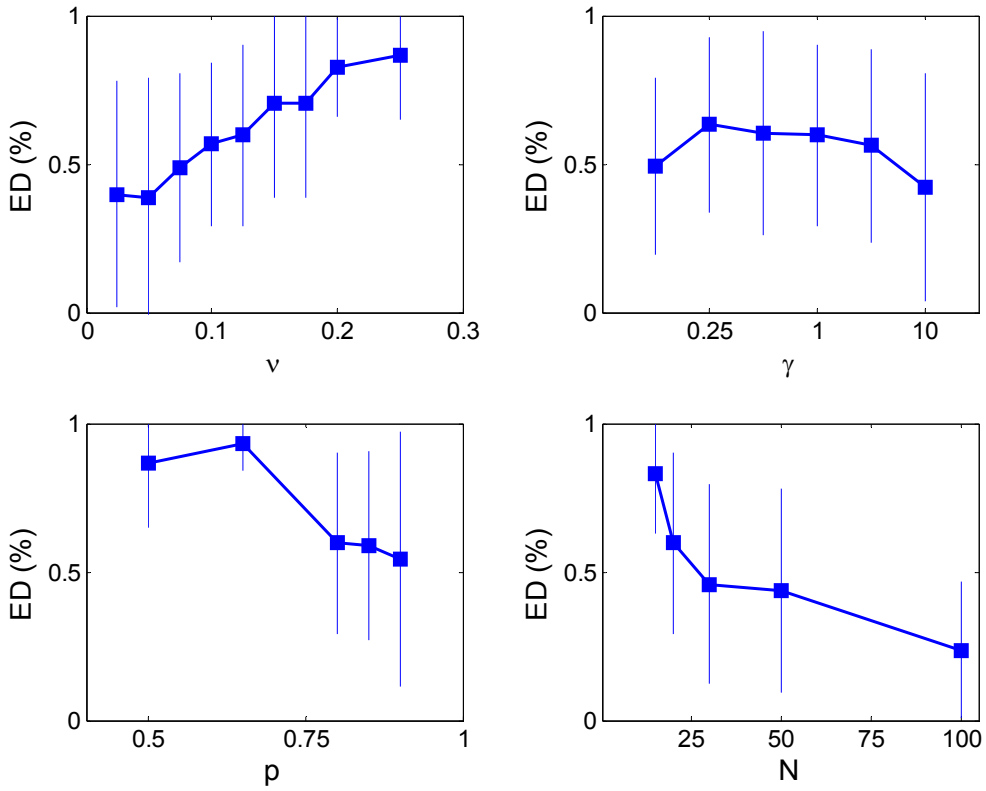

Fig. 4-16. Summary of parameter effects on estimated early detection efficacy of the novelty detector. Plots show mean early detection fraction values and 1-std error bars. Note that the abscissa for gamma is nonlinearly scaled for display purposes.

4.3.5 Other Effects

4.3.5.2 Detector Novelty State Distribution

During experiments it was observed that the estimated probability of the detector novelty state generally remained low, except during ictal events. This prompted an investigation of the distribution of the detector novelty state (Figure 4-17). The cumulative distribution function (*cdf*) was estimated using the WSTATS package. An

exponential distribution was selected for modeling. The maximum-likelihood estimated parameter was 0.1804; the estimated pdf of the exponential distribution for this data is

$$f(x) = \frac{\exp\left(-x/0.1804\right)}{0.1804} \tag{4.1}$$

Notice that the distribution of novelty state during ictal epochs deviates significantly from the baseline epochs. The flat region of the cdf shows an insensitivity in detector behavior between 0.35 and 0.75, suggesting that the novelty detection threshold should be set at $p \leq 0.35$ or $p \geq 0.75$.

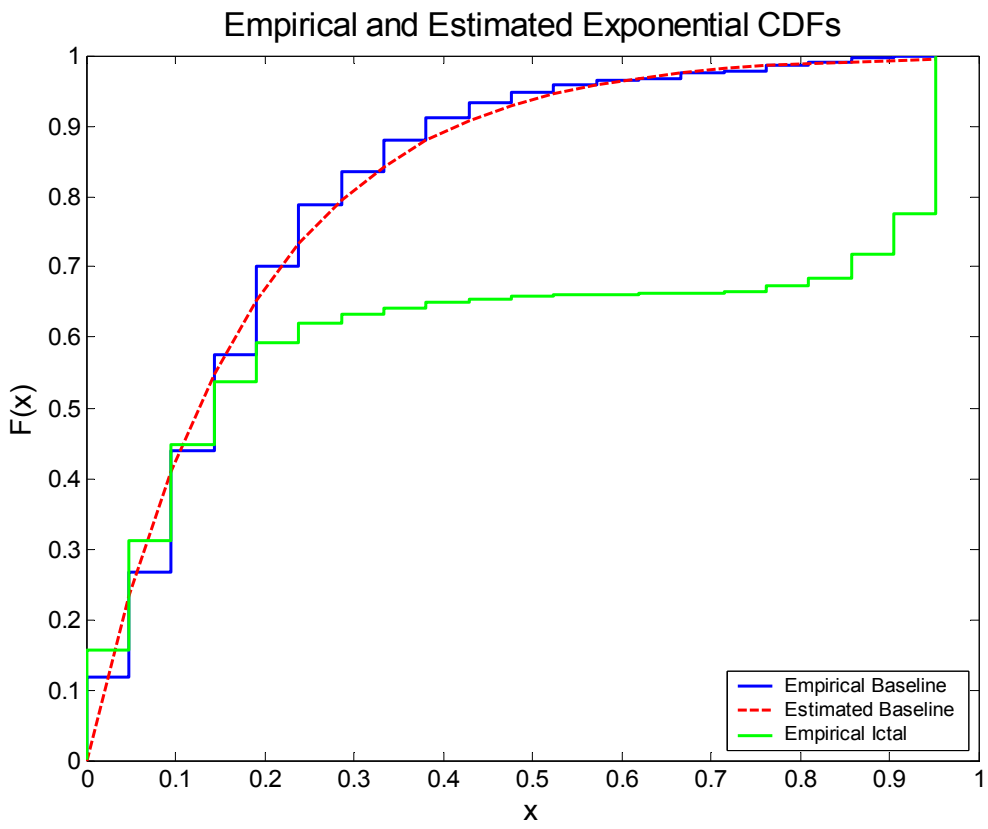

Fig. 4-17. The cumulative distribution of probability-estimates for three cases: empirical baseline segments, estimated baseline segments, and empirical ictal segments.

4.3.6 Performance Optimization

Optimal detector model parameters were determined by genetic algorithm (GA) optimization. The objective function for the optimization was heuristically defined to emphasize detector sensitivities of at least 75%, early-detection fractions, and small mean detection latencies, while penalizing false positive rates greater than 3.0 Fph. To describe the four optimization criteria quantitatively, four terms, k_1 to k_4, were defined:

$$k_1 = median(\mathbf{S}) \tag{4.2}$$

$$k_2 = median(\mathbf{EDF}) \tag{4.3}$$

$$k_3 = median(\mathbf{FPR}) \tag{4.4}$$

$$k_4 = median(\boldsymbol{\mu}_{\boldsymbol{\tau}}) \tag{4.5}$$

where $(\mathbf{S}, \mathbf{EDF}, \mathbf{FPR}, \boldsymbol{\mu})$ are vectors of performance statistics for each of the five patients. For example, $\mathbf{S} = (1,1,0.4,1,1)$, would represent sensitivities of 100% for Patients 1, 4, 7, 10, and a sensitivity of 40% for Patient 5. The median was selected for its robustness compared to the mean statistic.

Scoring of these four criteria was accomplished by equations 4.6 – 4.10. Scoring accounts for the relative importance of each performance metric. It is easy to interpret the scoring functions graphically (Figure 4-18). Clearly, the most important criteria is a high sensitivity (greater than 75%).

$$\alpha(k_1) = 100 \cdot k_1 - 10 \cdot (1 - \text{sgn}(k_1 - 0.75)) \tag{4.6}$$

$$\alpha(k_2) = 20 \cdot k_2 \tag{4.7}$$

$$\alpha\left(k_3\right) = -10 \cdot k_3 - 20 \cdot \left(1 - \mathrm{sgn}\left(5 - k_3\right)\right) \tag{4.8}$$

$$\alpha\left(k_4\right) = \max\left(-k_4, 30\right) \tag{4.9}$$

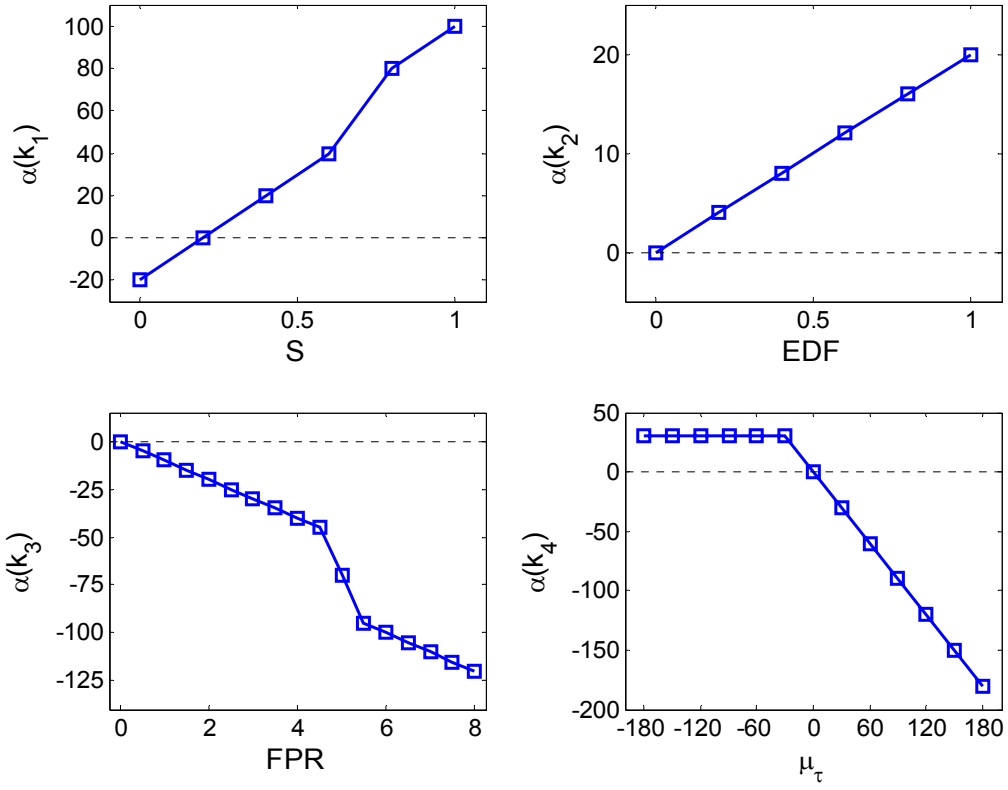

Fig. 4-19. The utility scores associated with each performance metric for use in the genetic algorithm optimization of the detector parameters. The utility scores are combined linearly in the objective function.

The resultant objective function is given in (4.10), and the range of parameters considered is given in 4.11.

$$\theta = \sum_{i=1}^{4} \alpha(k_i) \tag{4.10}$$

$$\begin{aligned} 0.2 \le \nu \le 0.2, \quad & 0.25 \le \gamma \le 10, \quad 0.3 \le p \le 1, \\ & 10 \le N \le 100, \quad 10 \le T \le 200 \end{aligned} \tag{4.11}$$

Valid scores range between -320 to 150. A "good" score, obtained by default parameters, is 80 or higher.

The Genetic Algorithm Optimization Toolbox (GAOT) was used for optimization. Floating-point chromosome values (as opposed to the conventional binary chromosomes) encoded the five model parameters. A population of 20 solutions was evolved for 40 generations. The best solution found scored 99.2: $\Pi = (\nu, \gamma, p, N, T) = (0.073, 0.3, 0.59, 38, 192)$.

4.3.7 Summary of Parameter Tuning

The effects of tuning of each model parameter were characterized. Several key design tradeoffs were identified as a result: for greater detector sensitivity, it is important to increase N and decrease ν; however, doing so also decreases the mean detection latency. Increasing the novelty detection threshold, p, decreases the FPR but results in fewer early detections and higher mean latencies. The particular choice of model parameters is application-dependent.

It is interesting to note that the FNR is non-zero when additional baseline segments are used for cross-validation. It is hypothesized that this is due to different

patient states-of-consciousness for the epochs in the dataset. This could also be due to the training data not fully representing the full spectrum of patient states, due to the relatively small sample size of data epochs, and their overall relatively small percentage of the entire data set. The FNR remains low over the range of parameters, so the detector sensitivity seems to remain quite high regardless of parameter settings.

Under the assumption of frame-independence, the binomial distribution can be used to estimate the probability that the novelty detection threshold, p, is exceeded:

$$P(\hat{p} > p) = 1 - F(n) = 1 - \sum_{i=0}^{n} \binom{N}{i} \nu^i (1-\nu)^{(N-i)} \tag{4.12}$$

which, for $\nu = 0.1$, is less than 0.002.

Clearly the empirical cdf for ictal epochs is inconsistent with this assumption, while baseline epochs are consistent with a binomial distribution. This can be interpreted in two ways: (1) feature vectors from ictal epochs are derived from a different distribution than those from baseline epochs, and/or (2) feature vectors from ictal segments are not independent. Both of these interpretations support the hypothesis that epileptiform activity is different from baseline brain activity (e.g., epileptiform activity is novel).

4.4 Summary of Static Detector Results

The static detector experiments provide an estimate of the performance of the system in an offline setting. The key results are: 100% sensitivity, -7.08 mean detection latency, and 41% early detection fraction at 1.57 Fph with a persistence of T = 180

seconds. This outperformed the state-of-the-art for all metrics except FPR. The possible presence of other (non-ictal anomalies), and the acceptance by the research community of hyperdetection strategies, diminishes the emphasis placed on FPR. For example, in early prototype reactive stimulation devices to treat seizures, the very brief and subthreshold stimulation involved (e.g. it does not cause afterdischarges) appears to be well tolerated without any significant side-effects, at this time. In this setting, the need to prevent seizures, (avoid false negative events) and the relative harmlessness of false positive stimulations, encourage making the detector "hypersensitive." The static detector successfully detected seizures without regard to state-of-consciousness and without prior knowledge or patient-specific tuning. The present method is computationally efficient due to the use of the SVM.

Several clear design tradeoffs exist. A larger value of detector persistence decreases the FPR, but increases the mean detection latency; a nominal value of T = 180 seconds was used. The sensitivity of the novelty detector depends solely on the FNR, which is in turn dependent on the model parameters, N and ν. Adjusting these parameters to increase sensitivity increases FPR as well. A similar tradeoff between early detection fraction and FPR, and mean detection latency and FPR exists. A genetic algorithm optimization using a heuristically derived objective function determined optimal detector parameters: $\Pi = (\nu, \gamma, p, N, T) = (0.073, 0.3, 0.59, 38, 192)$.

Chapter 5

Online Detector Results

This chapter presents the results of a naïve online novelty detection system as a seizure analysis tool. The naïve detector extends the static detector presented in Chapter 4 by implementing simple, fixed-schedule retraining of the 1-class SVM. This retraining allows the detector to adapt in a patient- and data-specific manner. Results from the static detector experiments guided selection of model parameters. The online detector was evaluated on the full 1077 hours of patient data in the epilepsy corpus (cf. Chapter 2). Patient-specific and summary results are presented, along with issues identified in implementing an online detector. The static and online detector results are compared. These findings indicate that the naïve online detector achieves state-of-the-art seizure detector performance.

5.1 Methods

The results of experiments with the static novelty detector prompted an implementation of an online algorithm. Optimal model parameters determined from static experiments (cf. 4.1.2) were employed. A simple adaptation strategy was implemented, and the detector was evaluated on focus channel data from the entire IEEG corpora. The performance of the online detector was evaluated using the same four key performance criteria previously identified.

5.1.1 Data Preparation

Data from a single channel was analyzed. Channel selection was patient-specific, and corresponded to the channel declared as the focus most frequently by the epileptologist. It is important to note that for a given patient, not all seizures necessarily originate in the same channel, nor do all seizures exhibit in all channels. The onset times marked by the epileptologist refer to the earliest seizure onset, and they are seizure- and channel specific. Onset times for seizures that propagate to additional channels are not marked. This affects performance assessment of the detector when the seizure origin is not the same as the channel analyzed. These occurrences are noted and discussed in the presentation of results. Most of these data were also considered by Echauz et al. [31] in an offline experiment. In total, more than 1077 hours of focus channel data (1.2 GB of raw data) was analyzed.

5.1.2 Online Detector Architecture

The system architecture for the naïve online novelty detector was shown in Figure 3-4. The system is composed of the same elements used by the static detector: a 1-class SVM for classifying 1-second frames as novel or normal, a probability estimation technique (the sequential hypothesis test) for estimating the probability that the detector is in the novel state, and a refractory period (persistence) to limit spurious detector outputs. The output of the novelty detection system is an indicator variable, $z[n]$, labeling IEEG intervals as novel or normal. The optimal model parameters (cf. 4.3.6) for the SVM, sequential hypothesis test, and persistence output were used, $\Pi = (\nu, \gamma, p, N, T) = (0.073, 0.3, 0.59, 38, 192)$.

The online nature of the detector requires an extension to the basic static novelty detector to allow for adaptation, or learning, over time. A simple adaptation technique is to perform a complete retraining of the 1-class SVM according to a retraining schedule. In this way the 1-class SVM is able to adapt to distributional changes in the IEEG feature vectors over time. Such changes stem from changes in patient state-of-consciousness, ictal events, and other phenomena. The naïve detector derives its name from the simple retraining strategy employed: the 1-class SVM is retrained every epoch on feature vectors derived from the previous epoch. The decision to train on 15-minute epochs was made based on the success of the static novelty detector experiments using 15-minute baseline epochs for training. Figure 5-1 illustrates the adaptation strategy employed.

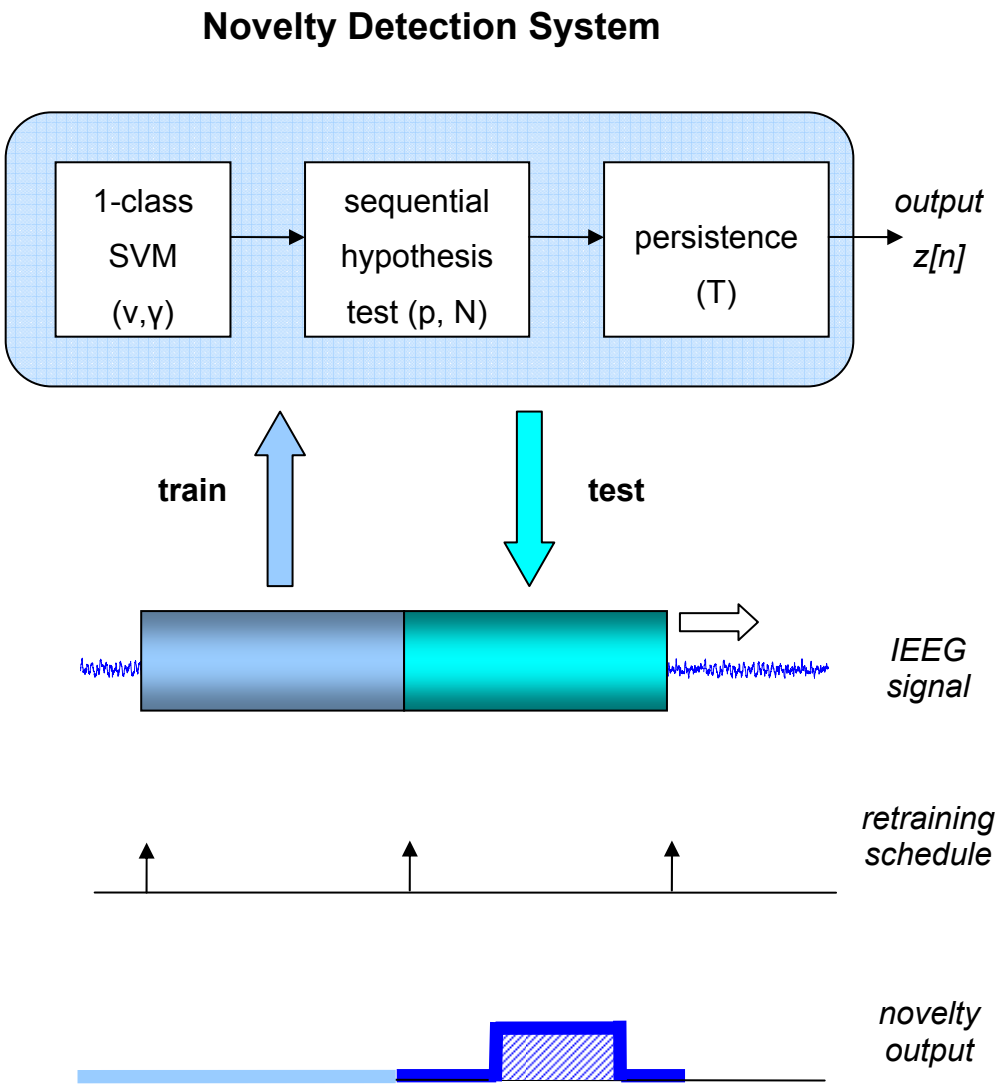

Fig. 5-1. Illustration of adaptation for the naïve online novelty detector for seizure analysis. Model parameters, $\Pi = (\nu, \gamma, p, N, T)$, remain fixed, but the SVM is retrained every 15-minutes on the previous 15-minute epoch. The system output indicates intervals of novel IEEG activity.

5.2 Results

The online novelty detector was run on the entire dataset described in 5.1.1 and performance metrics were computed. Novelty events were reviewed and compared to onset times to compute performance metrics. Detailed detection results are presented in Appendix A. Summary results, patient-specific comments, and a comparison between the static and online detector results are presented in the following sections.

5.2.1 Summary Results

A summary of seizure detection performance results, by patient, for the online detector is presented in Table 5-1.

Table 5-1. Summary of online detector performance results. Columns P1 – P16 show performance metrics by patient. Detailed detection results are available in Appendix A.

	P1	**P2**	**P3**	**P4**	**P5**	**P7**	**P8**	**P9**	**P10**	**P11**	**P15**	**P16**
S (%)	100	100	87.5	100	100	100	100	100	100	100	100	93.1
EDF (%)	40.0	100	37.5	0	80.0	14.3	33.3	63.6	40.0	80.0	50.0	18.5
τ (secs)	15.3	-61.2	-33.1	5.3	-1.6	2.21	-23.5	-5.5	-20.0	-37.0	-25.5	-8.0
FPR (Fph)	2.34	2.30	1.47	0.67	1.29	0.45	1.32	1.02	2.66	3.69	1.77	1.88

Of the 118 seizures identified, 93 were evaluated using the online novelty detector. The remaining 25 seizures were excluded from analysis for one of the following reasons: (1) the data was not available in the corpora, (2) the onset occurred too close to a recording epoch boundary to allow the SVM to be trained on 15-minutes of data, (3) the origin of the seizure did not correspond to the channel under analysis so a precise onset time was not available.

The distribution of early-detection fractions is shown in Figure 5-2.

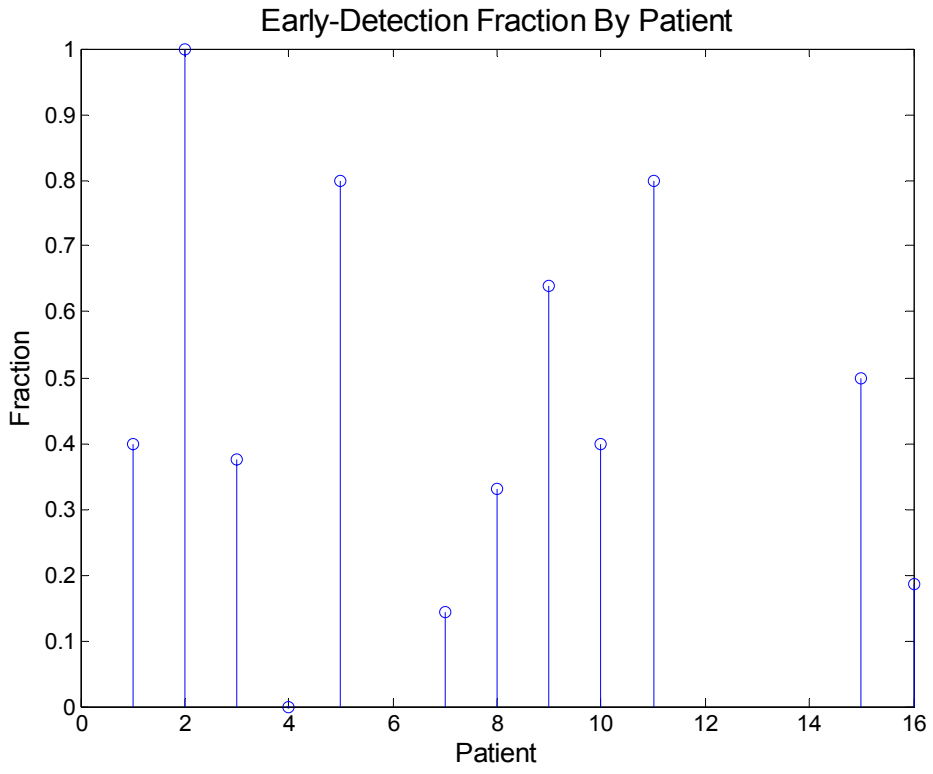

Fig. 5-2. Patient-specific summary of early-detection fraction.

37 seizures were early-detections resulting in an overall 40% early-detection fraction, consistent with that predicted by the static detector. Patients 2, 5, 9, and 11 seizures were mostly predicted. The mean and median detection latencies for seizures that were predicted were -48.4 and -20.0 seconds respectively. The overall mean and median of the detection latencies were -13.3 and 6.0 seconds, respectively. Figure 5-2 shows a histogram of detection latencies.

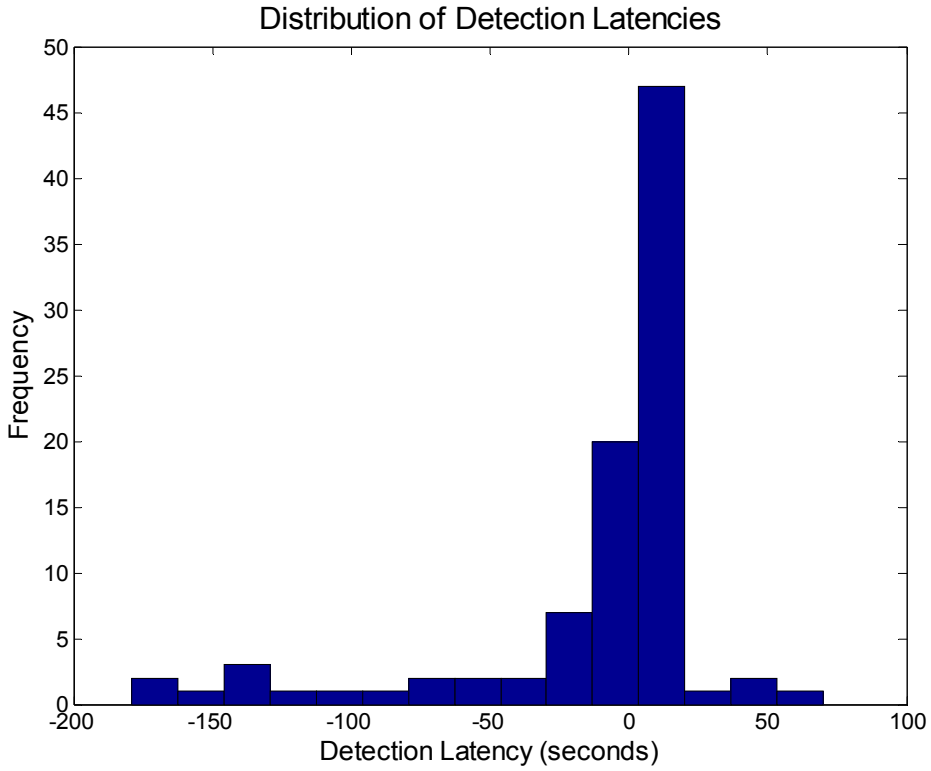

Fig. 5-3. Distribution of detection latencies for the naïve online novelty detector for seizure analysis.

The overall detector sensitivity was 97.85%. Only two of the seizures went undetected. These two seizures, however, occurred with 7-minutes of a preceding

seizure, so the detector did not undergo a retraining event. This may be the reason that the seizures were undetected.

The mean FPR across all patients was 1.74 Fph. The distribution of FPR by patient is shown in Figure 5-4.

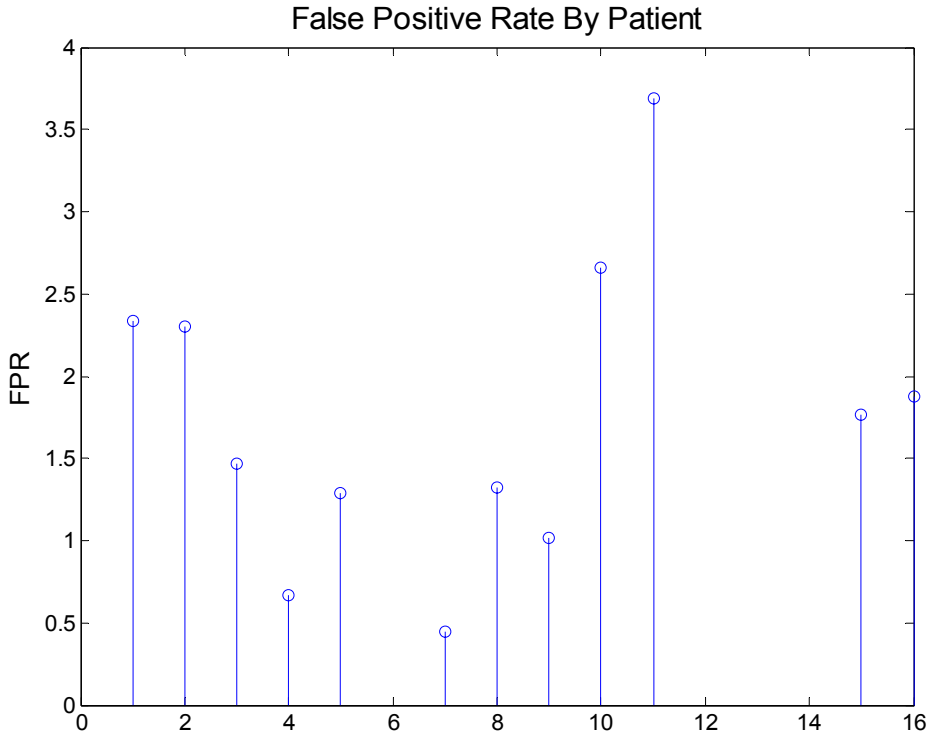

Fig. 5-4. Distribution of false positive rates for the naïve online novelty detector for seizure analysis.

5.2.2 Patient-Specific Comments

5.2.2.1 Patient 1

Seizure 2 occurred near the boundary of a recording epoch so the naïve online detector did not have 15-minutes of training data available. This seizure was excluded from the performance assessment calculations.

5.2.2.2 Patient 2

Seizure 2 was detected twice (the second detection had a latency of +8 seconds).

5.2.2.3 Patient 3

Seizures 7 and 8 were not available in the corpora and were excluded from performance assessment calculations. Epileptologist notes indicated that seizure 12 had a diffuse onset and contained reference artifacts in the IEEG recording.

5.2.2.4 Patient 4

Seizures 1 and 2 did not originate at the electrode location analyzed. Seizures 3 and 9 occurred within 15-minutes of a recording interval boundary, so the detector could not be trained prior to onset. Seizure 10 was not available in the corpora. Seizures 1, 2, 3, 9, and 10 were excluded from performance assessment calculations.

5.2.2.5 Patient 5

All seizures were analyzed successfully.

5.2.2.6 Patient 7

Seizures 0a, 0b, 5, and 6 did not originate at the electrode location analyzed; they were excluded from performance assessment calculations.

5.2.2.7 Patient 8

Seizure 3 did not originate at the electrode location analyzed; they were excluded from performance assessment calculations

5.2.2.8 Patient 9

All seizures were successfully analyzed.

5.2.2.9 Patient 10

Seizures 0a and 8 were not available in the corpora. Seizure 9 did not originate at the electrode location analyzed. Seizures 0a, 8, and 9 were excluded from performance assessment calculations.

5.2.2.10 Patient 11

All seizures were successfully analyzed.

5.2.2.11 Patient 15

Seizures 2, 3, 4, and 5 did not originate at the electrode location analyzed. Seizure 6 was not available in the corpora. Seizure 7 occurred within 15-minutes of a recording interval boundary, so the detector could not be trained prior to onset. Seizures 2 – 7 were excluded from performance assessment calculations.

5.2.2.12 Patient 16

Seizures 1d and 1ka occurred within 15-minutes of the previous seizure onset; this may explain why they were not detected. Seizures 4 and 7 did not originate at the electrode location analyzed; were excluded from performance assessment calculations. This patient was the only patient with false negative detections.

5.2.3 Discussion

A comparison of results between the static and online detectors is shown in Table 5-2.

Table 5-22. Comparison of performance results between the static and online detectors.

Metric	Static Detector	Naïve Online Detector
Sensitivity	100.0%	97.85%
Mean detection latency	-7.1-secs	-13.3-secs
Early-detection fraction	41.0%	40.0%
False positive rate	1.57 Fph	1.74 Fph

These results show that the static detector experiments provided good estimates of the performance of the online detector, and that the naïve online novelty detector also exceeds state-of-the-art performance at seizure detection

(except for FPR). The slight drop in sensitivity is easily explained by the disparity in the size of the datasets analyzed, and in the variance of estimating such a small fraction (the FNR). The mean detection latency improved to -13.3 seconds for the online detector. A close examination of the detector output reveals that this is attributable to seizures that were predicted. The early-detection fraction was very consistent. The false positive rate worsened slightly for the online detector, but this can be attributed to Patient 11's high FPR.

5.4 Performance Issues

This section discusses some key design and performance issues encountered when developing the naïve online novelty detector.

5.4.1 Retraining Event Schedule

Naïve scheduling is not the only possible rescheduling approach. Other approaches, especially dynamic approaches, should be investigated. While the fixed-schedule retraining is simple, it does not account for recent novelty activity. The novelty detector is retrained completely each epoch, regardless of the number of novel frames declared in the previous epoch. This has the effect of creating an overshoot phenomenon in the detector output where postictal epochs contain many novelty events. A more sophisticated retraining algorithm would exclude previous novelty frames from training. An alternative approach to retraining the online detector would be to update the support vectors of the SVM directly as each frame of data is processed. Adaptive SVMs are still an open research issue.

5.4.2 Probability Estimation

Probability estimation is a key factor in the success of the system. The probability estimation technique employed here is a simple; maximum likelihood estimate using a fixed window size. It is possible that more sophisticated probability estimation techniques could improve detector the performance. One such estimation technique that is effectively used for arithmetic compression is the Q-Coder algorithm [121]. The Q-Coder implements a fast, efficient renormalization strategy for probability estimation for an arithmetic coder. Such an approach is appealing for the novelty detector because of the binary output of the 1-class SVM used.

5.4.3 Effects of Persistence and Epoch Size

The effect of persistence is intuitive, but the interpretation of early-detections due to increased persistence values is unclear. Ten of the detected seizures had detection latencies of less than -60 seconds, and each of these were multiply detected. The subsequent detections typically occurred at +6.0 to +8.0 detection latency. The question arises: are these seizures really being predicted? It is not obvious how one might answer that question definitively given (1) such a small dataset of occurrences, and (2) an imprecise understanding of ictogenesis.

Epoch size is also important to the success of the system. No attempt was made o investigate the effects of epoch size on system performance, but this is an important computational consideration for implementation in a practical, implantable device. The smaller the epoch size, the faster the system adapts and the smaller the memory constraints of the device.

5.4.4 Interevent Duration Distribution

The distribution of inter-novelty events was examined and determined to be well-modeled by a lognormal distribution. Figure 5-5 shows the empirical and maximum-likelihood cdfs.

The estimated lognormal parameters are $\left(\mu = 3.37, \sigma^2 = 1.35\right)$. The pdf of the lognormal distribution is given as:

$$f\left(x\right) = \frac{1}{2\pi\left(\sigma x\right)^2} \exp\left\{\frac{\left(-\log x - m\right)^2}{2\sigma^2}\right\} \quad , x \geq 0 \qquad (5.1)$$

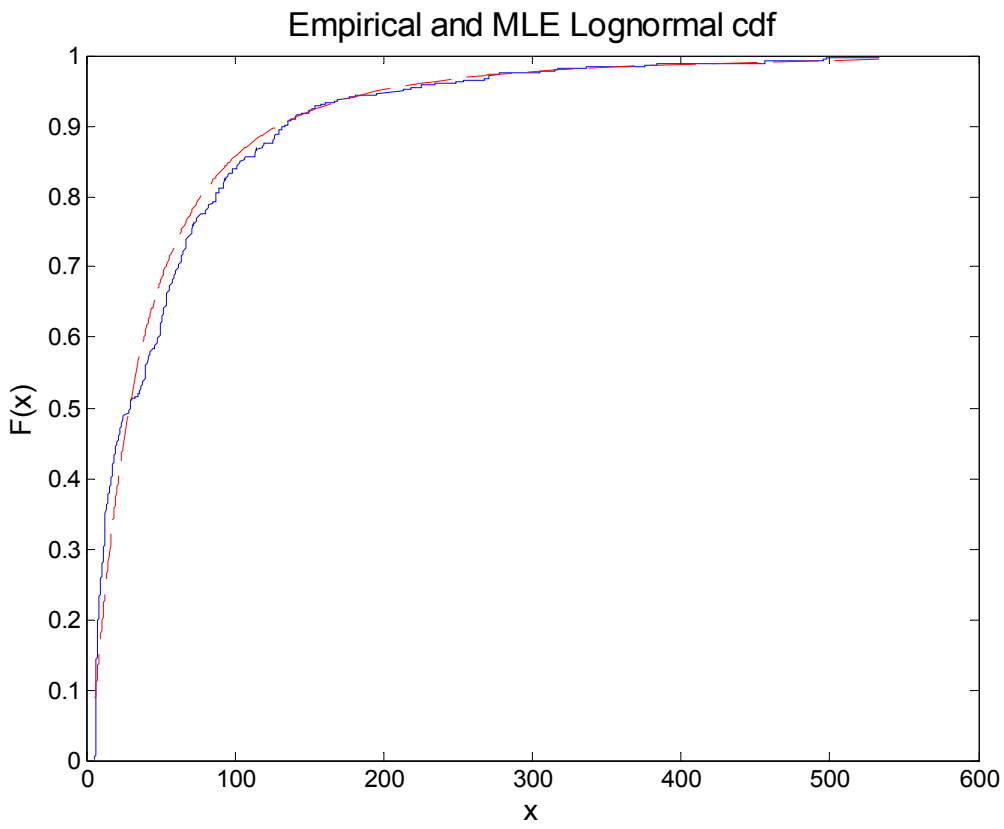

Fig. 5-5. The cumulative distribution function (cdf) of inter-event interval duration as approximated by the lognormal distribution.

Stochastic models of the novelty detection event process, and the seizure onset process, might be used to establish confidence bounds on detector performance.

Chapter 6

Other Experiments

This chapter discusses the results of preliminary experiments conducted using the novelty detection system presented in Chapter 3. The goal of these experiments was to explore extensions to the basic novelty detection paradigm for other seizure analysis functions, in particular for mapping the spread of seizure onset across multiple channels in the hope of automatically locating the focus channel. The findings from these experiments are presented in the following sections.

6.1 Multi-Patient, Multiple Channel Experiment

During development of the basic novelty detection system it was observed that there was a great degree of variability in detection latency results between channels. From this observation it was hypothesized that the basic novelty detector may be useful in mapping seizure onset, and possibly even identifying a focus channel automatically. This function would be of great value in practice as it is common for 64-, 128-, or 256-channels of data to be recorded simultaneously. Review of these data by human experts is time consuming, and in some cases a comprehensive review is not possible.

6.1.1 Methods

6.1.1.1 Data Preparation

The canonical focal dataset analyzed in Chapter 4 were analyzed for the experiments below.

6.1.1.3 Training

A static detector was trained (cf. 4.1.1.3) with default model parameters: $\Pi = (\nu, \gamma, p, N, T) = (0.1, 1, 0.8, 20, 180)$. Detector persistence (T = 180 seconds) was employed.

6.1.1.4 Testing

Testing was performed by analyzing the novelty detection output for each seizure segment, for each channel. The resulting detection latencies were recorded for each channel, seizure, and patient.

6.1.1.5 Channel Scoring

A mechanism for scoring the proximity of a channel to the focus was devised. In *channel scoring*, the channel ensemble means and standard deviations of latencies were computed for each patient. A scoring function was evaluated for each channel using the channel latency statistics and the ensemble statistics for the patient. The scoring function was designed to emphasize two criteria deemed relevant to focus channels: the ability of a channel (relative to other channels in the same patient) to provide early detection of seizures, and the low variance of the detection latencies associated with a channel. The first criterion is captured by comparing the mean latency from a channel to the ensemble

mean latency (Equation 6.1) and adopting a one-sided loss function. Figure 6-1 shows an example of loss functions.

$$L_k = \left(\mu_k - \frac{1}{N} \sum_{i=1}^{N} \mu_i \right) \tag{6.1}$$

where L_k is the basic scoring function, α_k is a coefficient penalizing high-variance channel results, k is the channel index, and the $\left(1 - \mathrm{sgn}\left(\cdot\right)\right)$ term makes the loss function positive.

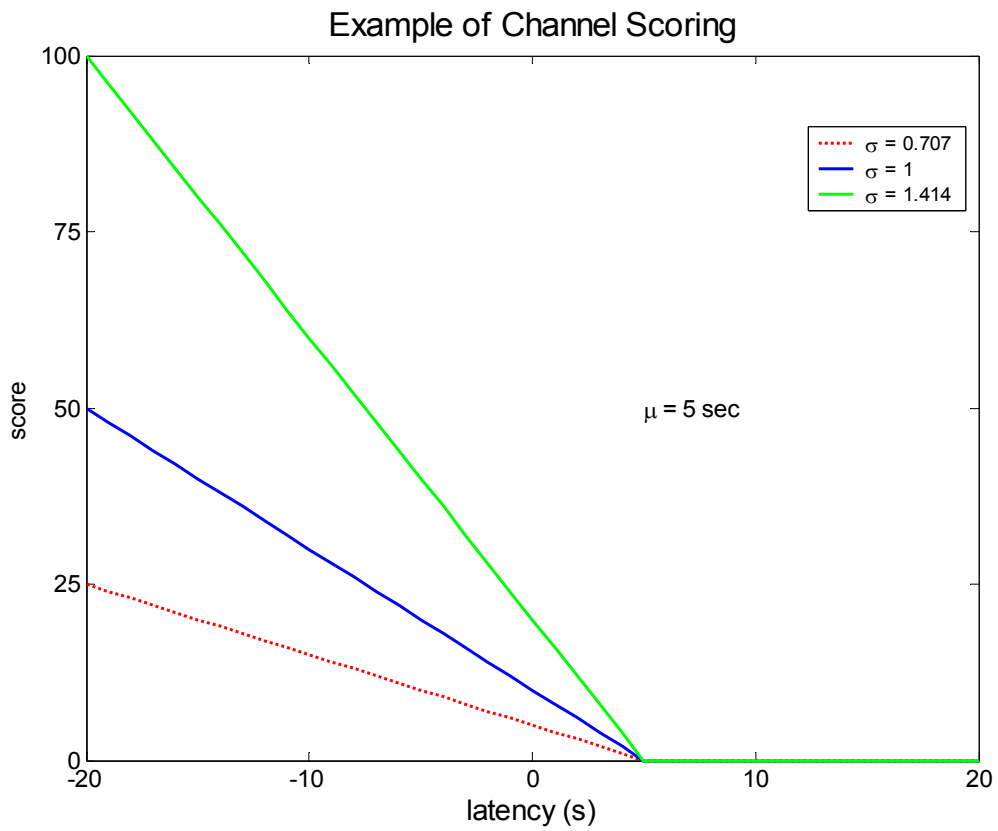

Fig. 6-1. The channel scoring function rewards better-than-average detection latency, penalizing for increased variance in the latency estimate. In this example three scoring functions are shown where the ensemble detection latency is 5 seconds.

The second criterion is captured by comparing the variances of the channel detection latencies to the ensemble detection latencies.

$$\alpha_k = \frac{\sigma_k}{\sum_{i=1}^{N} \sigma_i} \tag{6.2}$$

The final channel score, θ_k, is obtained as:

$$\theta_k = \alpha_k \left| L_k \right| \left(1 - \mathrm{sgn}\left(L_k \right) \right) \tag{6.3}$$

6.1.2 Results

The results of the per-channel analysis are presented graphically in the following sections (by patient), along with a brief discussion. For each patient, a figure is presented showing the mean detection latency by channel and associated 1-standard-deviation error bars. The per-channel scores are also shown. Summary results are discussed at the end of this section.

6.1.2.1 Patient 1

Patient 1 had six seizures; all but one of which were well-detected in all channels. The epileptologist marking this record indicated that the seizures appeared to originate in the left temporal lobe (LT 2-3 electrode contacts). This is consistent with the channel scoring result, as these two channels exhibit the highest channel score. This finding agrees with a general impression in the epilepsy community that focus channels are typically the best channels for early-detection of seizures, though there is no uniform agreement on this issue [4]. In their study of seizure prediction using a 10-minute prediction horizon, D'Alessandro et al. suggest that homologous contralateral channels may be better for seizure prediction than the focus region because the change from baseline to preictal state is of greater absolute magnitude than in the focus channel, where the change from abnormal baseline to seizure onset is less dramatic.

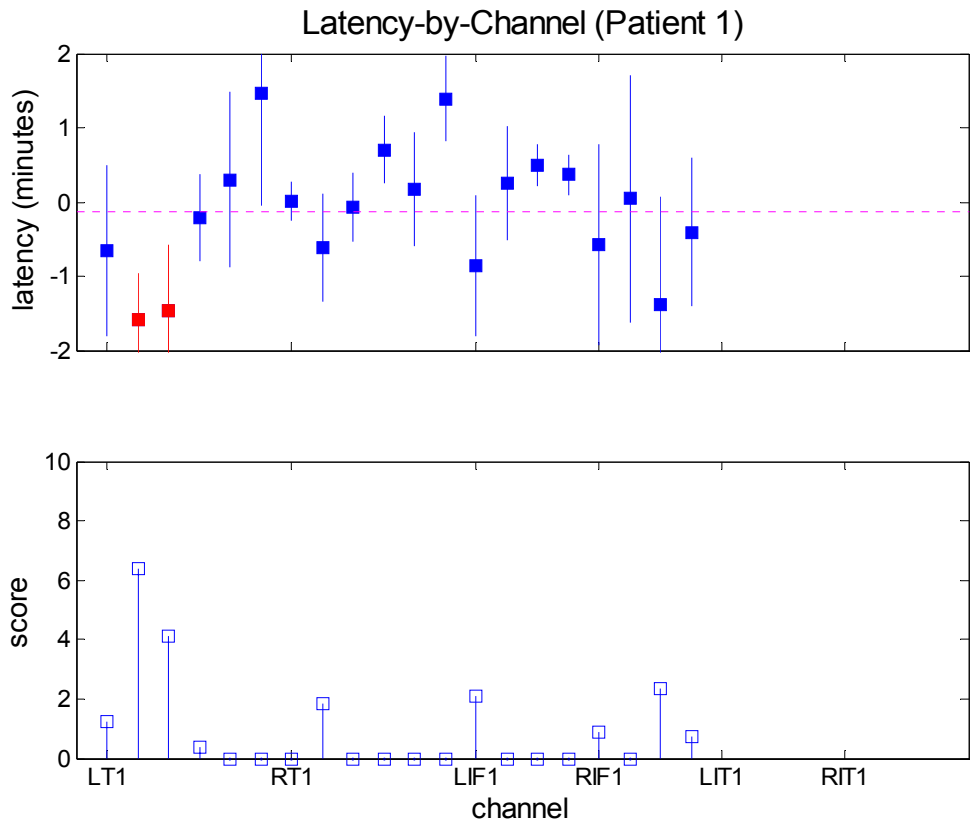

Fig. 6-2. Detection latencies by channel for Patient 1. A total of six seizures were analyzed. The focus channels are indicated in red, the ensemble mean in magenta. Channel scoring indicates that the focus channel is the best channel for early detection potential.

6.1.2.2 Patient 4

Patient 4 exhibited nine seizures, however two of the seizure records were not considered for analysis as the focus channel was not consistent with the remaining data. Seizures were well-detected across all channels and generally exhibited consistent detection as evidenced by the small variance of detection latency in many of the channels. The frontal lobes showed the most promise for early detection potential, however the right temporal lobe also indicated early onset. It should be noted that two

seizures were not detected in the RIF region. These results support the epileptologist marking of RT 2-4 as the focus channel. Note also that the frontal lobe regions exhibited consistently high variance in results.

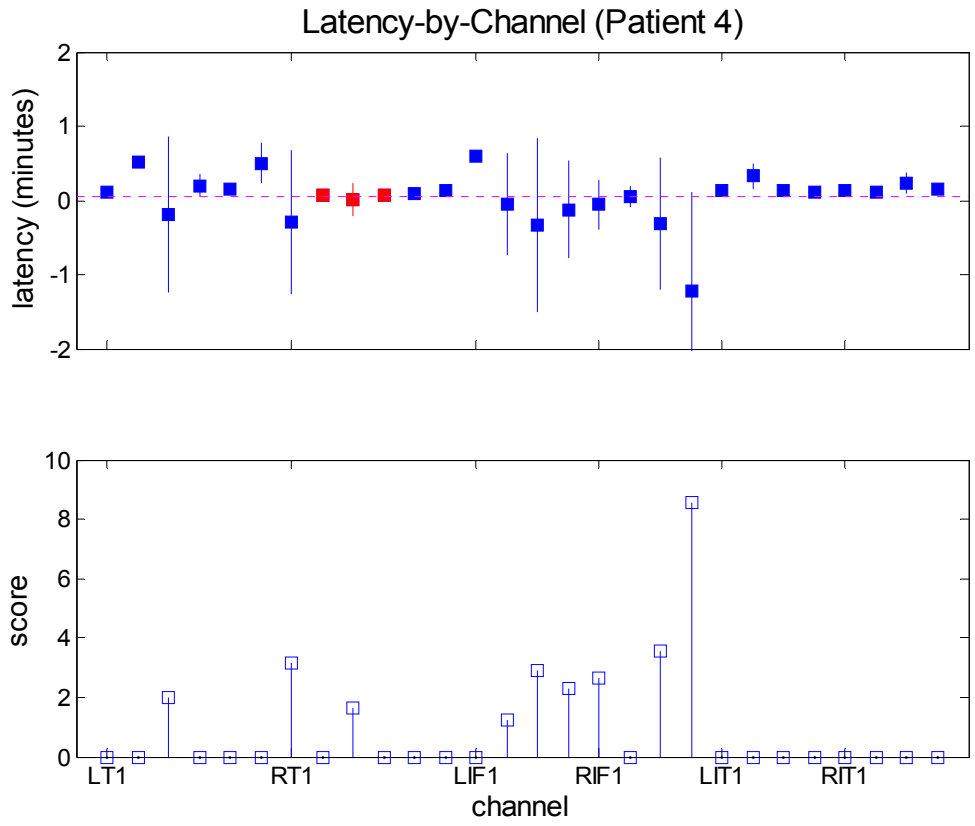

Fig. 6-3. Detection latencies by channel for Patient 4. A total of seven seizures were analyzed. The focus channels are indicated in red, the ensemble mean in magenta. Channel scoring indicates that the focus channel, as labeled by expert epileptologists, is good, but not best, for early detection potential. Alternatively, this result might indicate that expert readers may have not localized the seizure onset zone properly, and that this quantitative method might be a very useful tool for patient evaluation during evaluation for epilepsy surgery.

6.1.2.3 Patient 5

Patient 5 exhibited five well-stereotyped seizures. Seizures were well-detected across all channels with very little variance in results. Curiously, LIT3 was determined to be the best channel for early detection potential, however the channel scoring results support the epileptologist marking of RT 2-3 as the focus channel. All channels are very consistent at detecting seizures.

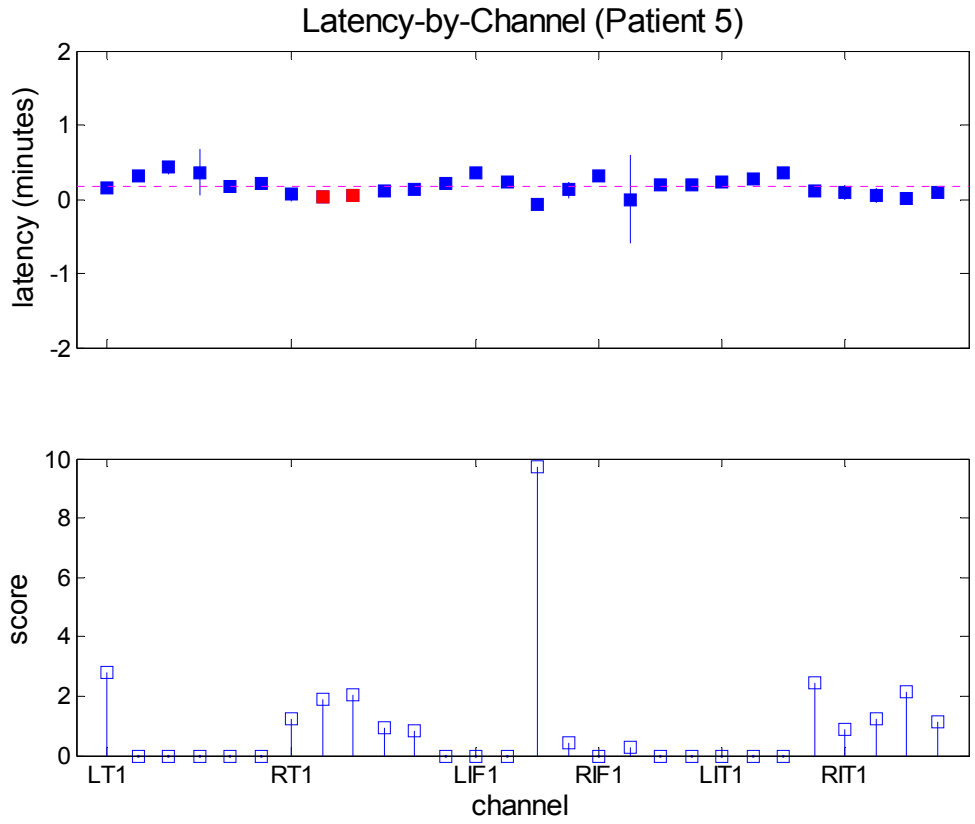

Fig. 6-4. Detection latencies by channel for Patient 5. A total of five seizures were analyzed. The focus channels are indicated in red, the ensemble mean in magenta. Channel scoring indicates that the focus channels are good (LIF3 is best) for early detection potential.

6.1.2.4 Patient 7 Results

Patient 7 exhibited 11 seizures. Nine of the seizures originating from the same focal region were examined. The detection latency results are highly-variable across most channels. LIF1 is indicated as the best early detection channel by channel scoring, however channel scoring results support the epileptologist marking of LT 3 as the focus channel.

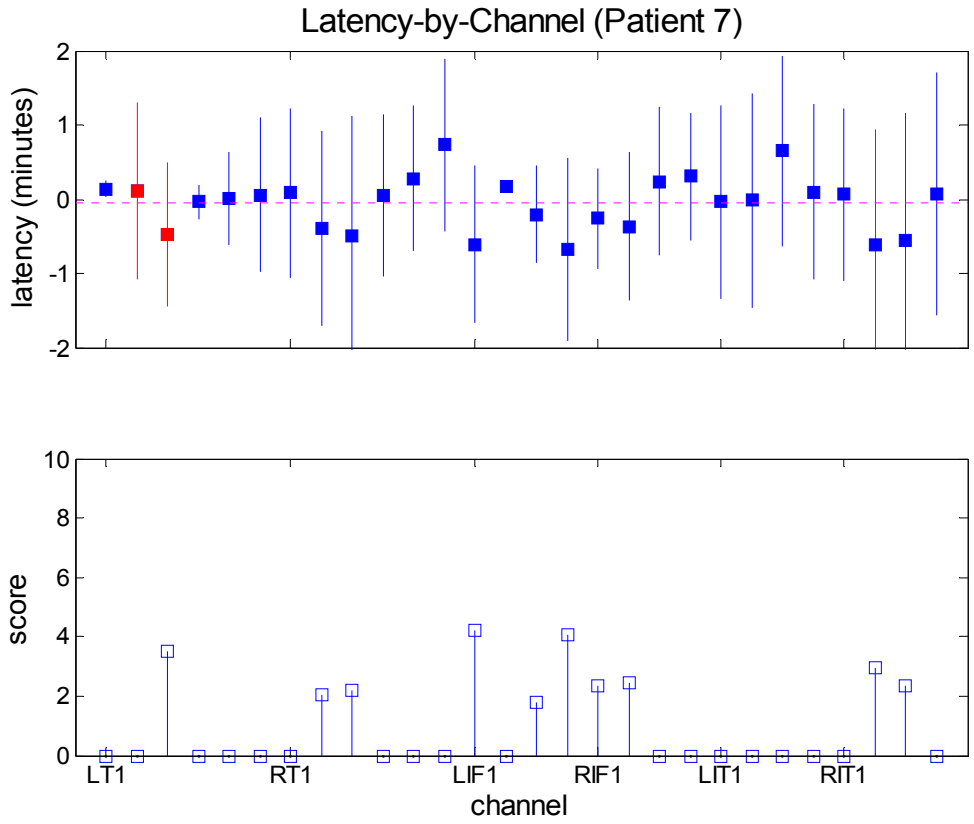

Fig. 6-5. Detection latencies by channel for Patient 7. A total of nine seizures were analyzed. The focus channels are indicated in red, the ensemble mean in magenta. Channel scoring indicates that one of the focus channels (LT3) is good for early detection potential, as are frontal lobe channels LIF1 and LIF4.

6.1.2.5 Patient 10 Results

Patient 10 exhibited 13 seizures. Ten of the seizures originating from the same focal region were examined. The detection latency results are highly-variable across most channels. Channel scoring indicates that the right-frontal and right-temporal regions are best for early detection potential. The channel scoring results do support the epileptologist marking of LT3 as the focus channel.

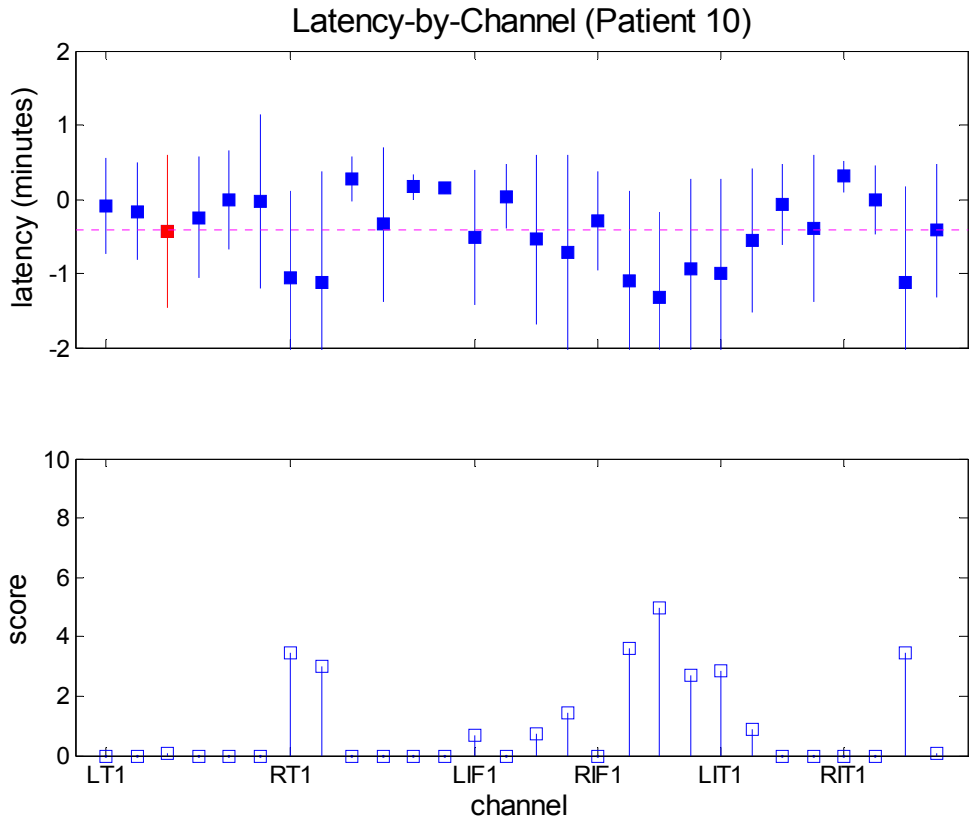

Fig. 6-6. Detection latencies by channel for Patient 10. A total of ten seizures were analyzed. The focus channels are indicated in red, the ensemble mean in magenta. Channel scoring indicates that RIF3 is best for early detection potential, and does not identify the focus (LT3) as a candidate channel.

6.1.3 Summary of Multi-Channel Results

The general observation from these experiments is that the novelty detector can provide a qualitative assessment of the onsets of seizures across patient channels. Derivation of a channel scoring function provides a good indicator of channels that are appropriate for seizure prediction or early-detection. Some patients exhibit very consistent latency results across all channels and seizures, while other patients, especially three with bilateral, independent seizure onsets, exhibit great variability in latencies. The potential for the channel scoring function and novelty detector per-channel latency assessment at providing automatic focus channel localization is promising. These results also support the findings by D'Alessandro [4] that seizure onsets may be better-predicted by examining non-focus channels.

Chapter 7

Conclusions and Future Research

This chapter summarizes the findings of this research, and offers direction for future research.

7.1 Conclusions

This research presents a new framework for analyzing time-series data and demonstrates its use for a practical problem: detecting and analyzing seizures from the intracranial EEG (seizure analysis). The fundamental hypothesis of the work—that seizures can be modeled as novelties in brain electrical activity and subsequently detected as such—was first verified on a small dataset. An online detection algorithm was devised, the naïve online novelty detector, and a comprehensive analysis was conducted on a large epilepsy database of continuous intracranial recordings, containing seizures and prolonged interictal periods, obtained from individuals with epilepsy. Finally, the novelty detector was successfully applied to track seizure onset and spread in multiple channels of intracranial EEG data.

By casting seizure analysis as a novelty detection problem, one eliminates the model bias of binary classification approaches, and leverages a robust statistical machine learning technique. The results of static and online experiments offer convincing proof that this technique has merit: the novelty detector outperformed the state-of-the art in seizure detection. Furthermore, the technique is computationally efficient, and requires

little patient training. The effects of model parameters important to the performance of the algorithm were investigated and an optimal parameter set was determined. Three key conclusions are drawn:

- Novelty detection is a successful, new technique for seizure analysis that outperforms the state-of-the-art [31] in seizure detection, achieving 97.85% detection sensitivity, 40% early-detection fraction, and -13.3 second mean detection latency at 1.74 Fph.
- A novelty detector can operate "out-of-the-box" with only 15-minutes of training data, no patient-specific training, and no knowledge of patient sleep state, using a simple adaptation strategy.
- Novelty detection can be applied to localize the onset of seizures in patients with focal epilepsies, perhaps outperforming expert epileptologists at this task.

7.2 Future Research

Future work can proceed in three directions: contributions to machine learning and signal processing, contributions to epilepsy research, contributions to other time-series applications.

The basic novelty detection architecture possesses theoretical properties that are not fully exploited in the body of work presented above. For instance, if one models the seizure onset process, perhaps as a Poisson or random walk stochastic process, theoretical aspects of the novelty detector, such as asymptotic novelty fraction, can be investigated.

These properties may provide more insight into model parameter settings, especially the relationship between persistence and performance metrics. It may be possible to establish a direct relationship between the statistical properties of the underlying time-series, e.g., variance, and optimal model parameters, e.g., ν (the novelty quantile parameter).

The online novelty detector employs a simple scheme for adaptation. Future research should compare scheduled-retraining of the 1-class SVM with direct-update of the support vectors of the SVM. For example, if the state of the novelty detector (e.g., the estimated probability that a novelty has occurred) were considered during retraining, it might be possible to avoid the "detector oscillations" observed in the experiments presented above. Two other extensions to the online novelty detector stand out as possible improvements: (1) use of hysteresis in conjunction with persistence for improved performance, (2) classification of novelty events that are detected to reduce false positives. For example, a hysteresis on the output of the persistence may allow for faster recovery after firing, and classifying novelty events prior to detector firing has the potential to drastically reduce false positive detections. Additionally, expert knowledge of motion artifacts—a generator of novelty events—could be used to prevent the detector from firing when a patient electrode moves.

Potential applications of the novelty detector to epilepsy research are numerous. In addition to obvious seizure or seizure precursor detection, it is possible that the preliminary results from Chapter 6 can be extended to develop a robust tool for automatic epileptic focus localization. This application could potentially help elucidate mechanisms underlying seizure spread in the epileptic network, as well as raise intriguing

questions. For example, it is possible that the algorithm may pick up changes in multiple channels, not obvious to the eye, that localize epilepsies previously considered to be nonfocal. Contributions to the buildup of activity prior to EEG onset of seizures may be found to involve channels located away from the focus, previously thought to be unimportant for seizure generation. Additionally, this technique might be used to standardize EEG markings by experts and aid in managing the enormous datasets they generate. Any reduction in epileptologist labor required to mark the very large datasets used for this type of research benefits the entire research community, and has great potential to improve researcher productivity. Examination of novelty frames of IEEG data may answer questions like "why did the detector fire in this case" during evaluation for closed loop stimulation devices, and may provide insight into the process of seizure generation (ictogenesis).

The role of individual features was not emphasized in this research, but it is clear that selection of appropriate features for detecting seizures was critical for the success of the novelty detector. The link between feature selection and system performance in classification is a central problem in machine learning. To address the issue of more general application of the novelty detection techniques refined in this work, the question arose, "Can the novelty detector operate as a feature selector?" Using the technique in this way could open up its application to problems in which useful features are not known. A simple scheme for implementing this application might "score" candidate features based on the performance results of the detector using those features. Combined with a technique for exploring candidate features, e.g., evolutionary algorithms, it may be possible to dramatically improve the performance of this system, even when good

candidate features are already known. Even further heuristic feature specifications may offer improvement: it is well-known, for instance, that time-frequency features show particular merit for seizure analysis.

The basic seizure detection application presented in this research operates on a single channel of data. Researchers have recently begun to question how systems incorporating multi-channel data can be devised. A straightforward extension of the novelty detector is possible by considering each data channel independently. This could be implemented using a mixture-of-experts technique from machine learning to produce a final labeling of novelty date, e.g., a voting scheme might operate on several channel outputs to simultaneously to produce a seizure declaration. In this regard, the output of the novelty detector can be viewed as itself being a feature for further processing.

While the research potential for the work begun in this dissertation appears great, one of the motivations for this research is its potential practical use in real-world applications. Specifically, the tremendous need for new treatments for epilepsy, coupled with the renaissance in devices and brain stimulation to treat neurological disease, have provided great inspiration. While further application of the above work is planned in a variety of intelligent systems in industry, defense, and maufacturing, perhaps the most satisfying application of these techniques will be their use in an implantable therapeutical neurological device.

Appendix A

This appendix contains the detailed seizure detection results obtained with the naïve online novelty detector. Results are tabulated for each patient. If a seizure went undetected (e.g., a false negative), the latency is marked as "X." Latencies marked with "—" indicate that the seizure was not considered for analysis for one of three reasons: (1) the data was missing, (2) the seizure occurred within 15-minutes of the beginning of an epoch, or (3) the seizure onset did not occur in the channel analyzed.

Table A-1. Summary of Patient 1 online seizure detection results.

Seizure	Latency
1	+ 19.0
2	--
3	+ 41.5
4	- 2.5
5	- 26.0
6	+ 44.5

Table A-2. Summary of Patient 2 online seizure detection results.

Seizure	Latency
1	- 20.0
2	- 146.0
3	- 93.0
4	- 35.0
5	- 12.0

Table A-3. Summary of Patient 3 online seizure detection results.

Seizure	Latency
1	- 2.0
7	--
8	--
9	+ 70.0
10	-179.0
11	+ 9.0
12	--
13	+ 8.0
14	+ 7.3
14a	-145.0

Table A-4. Summary of Patient 4 online seizure detection results.

Seizure	Latency
1	--
2	--
3	--
4	+ 6.5
5	+ 5.5
6	+ 1.5
7	+ 6.0
8	+ 7.0
9	--
10	--

Table A-5. Summary of Patient 5 online seizure detection results.

Seizure	Latency
1	- 4.5
2	- 0.5
3	+ 1.0
4	- 2.0
5	- 2.0

Table A-6. Summary of Patient 7 online seizure detection results.

Seizure	Latency
0a	--
0b	--
1	- 20.0
2	+ 9.5
3	+ 7.5
4	+ 0.5
5	--
6	--
7	+ 7.0
8	+ 3.5
9	+ 7.5

.

Table A-7. Summary of Patient 8 online seizure detection results.

Seizure	Latency
1	+ 5.0
2	+ 7.0
3	+ 28.0
4	-134.0

Table A-8. Summary of Patient 9 online seizure detection results.

Seizure	Latency
1	- 6.0
2	+ 6.5
3	+ 7.5
4	- 18.0
5	- 8.0
6	- 9.0
7	+ 3.5
8	- 12.0
9	+ 7.5
10	- 19.5
11	- 13.0

Table A-9. Summary of Patient 10 online seizure detection results.

Seizure	Latency
0a	--
1	+ 9.0
2	- 6.5
3	+ 7.0
4	- 22.5
5	+ 11.0
6	- 96.5
7	+ 8.0
8	--
9	--
9a	-126.5
9b	+ 9.5
10	+ 7.5

Table A-10. Summary of Patient 11 online seizure detection results.

Seizure	Latency
1	-145.0
2	6.0
3	- 10.0
4	- 2.0
5	- 33.8

Table A-11. Summary of Patient 15 online seizure detection results.

Seizure	Latency
1	+ 8.0
1a	- 59.0
2	--
3	--
4	--
5	--
6	--
7	--

Table A-12. Summary of Patient 16 online seizure detection results.

Seizure	Latency
1	+ 9.0
1aa	+ 8.0
1ab	+ 8.0
1ac	+ 7.0
1ad	+ 7.0
1a	+ 7.0
1b	+ 6.0
1c	+ 6.0
1d	X
1da	+ 7.0
1e	+ 8.0
1f	- 78.0
1g	+ 8.0
1h	+ 7.0
1i	+ 7.0
1j	+ 8.0
1k	- 78.0
1ka	X
1l	+ 8.0
1m	+ 8.0
1n	+ 8.0
2	- 1.0
3	+ 7.0
4	--
5	- 46.3
6	-176.0
7	--
8	+ 8.0
9	+ 8.0
10	+ 7.0
11	+ 6.0

References

[1] E. R. Kandel, J. H. Schwartz, and T. M. Jessel, *Principles of Neural Science*. New Jersey: Prentice-Hall, 1991.

[2] R. Esteller, "Detection of Seizure Onset in Epileptic Patients from Intracranial EEG Signals," in *ECE*. Atlanta: Georgia Institute of Technology, 2000.

[3] J. Gotman, "Seizure recognition and analysis," *Electroencephalogr Clin Neurophysiol Suppl*, vol. 37, pp. 133-45, 1985.

[4] M. D'Alessandro, "The Utility of Intracranial EEG Feature and Channel Synergy for Evaluating the Spatial and Temporal Behavior of Seizure Precursors," in *ECE*. Atlanta: Georgia Institute of Technology, 2001.

[5] B. Litt and J. Echauz, "Prediction of epileptic seizures," *The Lancet Neurology*, vol. 1, pp. 22 - 30, 2002.

[6] S. Liss, "Apparatus for monitoring and counteracting excess brain electrical energy to prevent epileptic seizures and the like." US, 1973.

[7] S. S. Viglione, V. A. Ordon, and F. Risch, "A methodology for detecting ongoing changes in the EEG prior to clinical seizures," presented at 21st Western Institute on Epilepsy, West Huntington Beach, CA, 1970.

[8] J. Gotman, "Automatic seizure detection: improvements and evaluation," *Electroencephalogr Clin Neurophysiol*, vol. 76, pp. 317-24, 1990.

[9] H. Qu and J. Gotman, "Improvement in seizure detection performance by automatic adaptation to the EEG of each patient," *Electroencephalogr Clin Neurophysiol*, vol. 86, pp. 79-87, 1993.

[10] H. Qu and J. Gotman, "A seizure warning system for long-term epilepsy monitoring," *Neurology*, vol. 45, pp. 2250-4, 1995.

[11] H. Qu and J. Gotman, "A patient-specific algorithm for the detection of seizure onset in long-term EEG monitoring: possible use as a warning device," *IEEE Trans Biomed Eng*, vol. 44, pp. 115-22, 1997.

[12] M. Zijlmans, D. Flanagan, and J. Gotman, "Heart rate changes and ECG abnormalities during epileptic seizures: prevalence and definition of an objective clinical sign," *Epilepsia*, vol. 43, pp. 847-54, 2002.

[13] H. R. Theodore and L. F. Haas, "A personal computer controlled long-term video-EEG monitoring, seizure detection, and collection system," *American Journal of EEG Technology*, vol. 31, pp. 103-108, 1991.

[14] H. P. Zaveri, W. J. Williams, and J. C. Sackellares, "Energy based detection of seizures," presented at 15th Annual International Conference on Engineering and Medicine in Biology, 1993.

[15] S. J. Schiff, J. Heller, S. R. Weinstein, and J. G. Milton, "Controlled wavelet transforms for EEG spike and seizure localization," *PRoceedings of the SPIE*, vol. 2242, pp. 762-775, 1994.

[16] E. S. Hughes, W. D. Gaillard, G. M. Jacyna, and S. J. Schiff, "Feature-based detection of epileptic seizures using ECoG recordings and a PAM," presented at

Proceedings of 17th International Conference of the Engineering in Medicine and Biology Society, Montreal, Que., Canada, 1997.
[17] W. J. Williams, H. P. Zaveri, and J. C. Sackellares, "Time-Frequency Analysis of Electrophysiology Signals in Epilepsy," *IEEE Engineering in Medicine and Biology*, pp. 133-143, 1995.
[18] J. J. Benedetto and D. Colella, "Wavelet analysis of spectrogram seizure chirps," presented at Wavelet Applications in Signal and Image Processing III, 1995.
[19] J. W. Creekmore, J. Conry, D. Colella, and S. J. Schiff, "Matched-filter based detection and localization of epileptic seizures from spectrographic chirps in ECoG recordings," presented at Proceedings of 17th International Conference of the Engineering in Medicine and Biology Society, 1997.
[20] H. P. Zaveri, R. B. Duckrow, N. C. de Lanerolle, and S. S. Spencer, "Distinguishing subtypes of temporal lobe epilepsy with background hippocampal activity," *Epilepsia*, vol. 42, pp. 725-30, 2001.
[21] M. Sun, L.-S. Pon, M. L. Scheurer, and R. J. Sclabassi, "Application of time-frequency and wavelet analysis to the diagnosis of epilepsy," presented at Wavelet Applications VII, Orlando, FL, USA, 2000.
[22] A. J. Gabor, R. R. Leach, and F. U. Dowla, "Automated seizure detection using a self-organizing neural network," *Electroencephalogr Clin Neurophysiol*, vol. 99, pp. 257-66, 1996.
[23] A. J. Gabor, "Seizure detection using a self-organizing neural network: validation and comparison with other detection strategies," *Electroencephalogr Clin Neurophysiol*, vol. 107, pp. 27-32, 1998.
[24] B. Marchesi, A. L. Stelle, and H. S. Lopes, "Detection of epileptic events using genetic programming," presented at Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA, 1997.
[25] I. Osorio, M. G. Frei, and S. B. Wilkinson, "Real-time automated detection and quantitative analysis of seizures and short-term prediction of clinical onset," *Epilepsia*, vol. 39, pp. 615-27, 1998.
[26] I. Osorio, M. G. Frei, B. F. Manly, S. Sunderam, N. C. Bhavaraju, and S. B. Wilkinson, "An introduction to contingent (closed-loop) brain electrical stimulation for seizure blockage, to ultra-short-term clinical trials, and to multidimensional statistical analysis of therapeutic efficacy," *J Clin Neurophysiol*, vol. 18, pp. 533-44, 2001.
[27] I. Osorio, M. G. Frei, J. Giftakis, T. Peters, J. Ingram, M. Turnbull, M. Herzog, M. T. Rise, S. Schaffner, R. A. Wennberg, T. S. Walczak, M. W. Risinger, and C. Ajmone-Marsan, "Performance reassessment of a real-time seizure-detection algorithm on long ECoG series," *Epilepsia*, vol. 43, pp. 1522-35, 2002.
[28] R. Esteller, G. Vachtsevanos, J. Echauz, M. D'Alessandro, C. Bowen, and R. Shor, "Fractal dimension detects seizure onset in mesial temporal lobe epilepsy," presented at Proceedings of the First Joint BMES/EMBS Conference, Atlanta, GA, USA, 1999.
[29] R. Esteller, G. Vachtsevanos, J. Echauz, T. Henry, P. Pennell, C. Epstein, R. Bakay, C. Bowen, and B. Litt, "Fractal dimension characterizes seizure onset in

epileptic patients," presented at 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing, Phoenix, AZ, USA, 1999.
[30] R. Esteller, J. Echauz, T. Tcheng, B. Litt, and B. Pless, "Line length: an efficient feature for seizure onset detection," presented at Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2001.
[31] J. Echauz, "Long-Term Validation of Detection Algorithms Suitable for an Implantable Device," A. B. Gardner, Ed., 2001.
[32] M.-J. Hoeve, R. D. Jones, G. J. Carroll, and H. Goelz, "Automated detection of epileptic seizures in the EEG," presented at 2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering n Medicine and Biology Society, Istanbul, Turkey, 2001.
[33] P. Celka and P. Colditz, "A computer-aided detection of EEG seizures in infants: a singular-spectrum approach and performance comparison," *IEEE Trans Biomed Eng*, vol. 49, pp. 455-62, 2002.
[34] P. E. McSharry, T. He, L. A. Smith, and L. Tarassenko, "Linear and non-linear methods for automatic seizure detection in scalp electro-encephalogram recordings," *Med Biol Eng Comput*, vol. 40, pp. 447-61, 2002.
[35] C. Baumgartner, W. Serles, F. Leutmezer, E. Pataraia, S. Aull, T. Czech, U. Pietrzyk, A. Relic, and I. Podreka, "Preictal SPECT in temporal lobe epilepsy: regional cerebral blood flow is increased prior to electroencephalography-seizure onset," *J Nucl Med*, vol. 39, pp. 978-82, 1998.
[36] V. V. Novak, L. A. Reeves, P. Novak, A. P. Low, and W. F. Sharbrough, "Time-frequency mapping of R-R interval during complex partial seizures of temporal lobe origin," *J Auton Nerv Syst*, vol. 77, pp. 195-202, 1999.
[37] V. Novak, A. L. Reeves, P. Novak, P. A. Low, and F. W. Sharbrough, "Time-frequency mapping of R-R interval during complex partial seizures of temporal lobe origin," *J Auton Nerv Syst*, vol. 77, pp. 195-202, 1999.
[38] R. S. Delamont, P. O. Julu, and G. A. Jamal, "Changes in a measure of cardiac vagal activity before and after epileptic seizures," *Epilepsy Res*, vol. 35, pp. 87-94, 1999.
[39] C. Schernthaner, G. Lindinger, K. Potzelberger, K. Zeiler, and C. Baumgartner, "Autonomic epilepsy--the influence of epileptic discharges on heart rate and rhythm," *Wien Klin Wochenschr*, vol. 111, pp. 392-401, 1999.
[40] P. Rajna, B. Clemens, E. Csibri, E. Dobos, A. Geregely, M. Gottschal, I. Gyorgy, A. Horvath, F. Horvath, L. Mezofi, I. Velkey, J. Veres, and E. Wagner, "Hungarian multicentre epidemiologic study of the warning and initial symptoms (prodrome, aura) of epileptic seizures," *Seizure*, vol. 6, pp. 361-8, 1997.
[41] V. Strong, S. W. Brown, and R. Walker, "Seizure-alert dogs--fact or fiction?," *Seizure*, vol. 8, pp. 62-5, 1999.
[42] S. W. Brown and V. Strong, "The use of seizure-alert dogs," *Seizure*, vol. 10, pp. 39-41, 2001.
[43] D. J. Dalziel, B. M. Uthman, S. P. McGorray, and R. L. Reep, "Seizure-alert dogs: a review and preliminary study," *Seizure*, vol. 12, pp. 115-120, 2003.
[44] V. Strong, S. Brown, M. Huyton, and H. Coyle, "Effect of trained Seizure Alert Dogs on frequency of tonic-clonic seizures," *Seizure*, vol. 11, pp. 402-5, 2002.

[45] J. Gotman and D. J. Koffler, "Interictal spiking increases after seizures but does not after decrease in medication," *Electroencephalogr Clin Neurophysiol*, vol. 72, pp. 7-15, 1989.
[46] I. Sherwin, "Interictal--ictal transition in the feline penicillin epileptogenic focus," *Electroencephalogr Clin Neurophysiol*, vol. 45, pp. 525-34, 1978.
[47] H. Lange, T. Tanaka, and R. Naquet, "Temporo-spatial pattern of subcortical spike activity in kindling epilepsy. A statistical approach," *Electroencephalogr Clin Neurophysiol*, vol. 42, pp. 564-74, 1977.
[48] M. Balish, P. S. Albert, and W. H. Theodore, "Seizure frequency in intractable partial epilepsy: a statistical analysis," *Epilepsia*, vol. 32, pp. 642-9, 1991.
[49] J. G. Milton, J. Gotman, G. M. Remillard, and F. Andermann, "Timing of seizure recurrence in adult epileptic patients: a statistical analysis," *Epilepsia*, vol. 28, pp. 471-8, 1987.
[50] B. Litt, R. Esteller, J. Echauz, M. D'Alessandro, R. Shor, T. Henry, P. Pennell, C. Epstein, R. Bakay, M. Dichter, and G. Vachtsevanos, "Epileptic seizures may begin hours in advance of clinical onset: a report of five patients," *Neuron*, vol. 30, pp. 51-64, 2001.
[51] M. C. Casdagli, L. D. Iasemidis, R. S. Savit, R. L. Gilmore, S. N. Roper, and J. C. Sackellares, "Non-linearity in invasive EEG recordings from patients with temporal lobe epilepsy," *Electroencephalogr Clin Neurophysiol*, vol. 102, pp. 98-105, 1997.
[52] L. D. Iasemidis, L. D. Olson, R. S. Savit, and J. C. Sackellares, "Time dependencies in the occurrences of epileptic seizures," *Epilepsy Res*, vol. 17, pp. 81-94, 1994.
[53] H. P. Zaveri, W. J. Williams, L. D. Iasemidis, and J. C. Sackellares, "Time-frequency representation of electrocorticograms in temporal lobe epilepsy," *IEEE Trans Biomed Eng*, vol. 39, pp. 502-9, 1992.
[54] L. D. Iasemidis, J. C. Sackellares, H. P. Zaveri, and W. J. Williams, "Phase space topography and the Lyapunov exponent of electrocorticograms in partial seizures," *Brain Topogr*, vol. 2, pp. 187-201, 1990.
[55] L. D. Iasemidis, H. P. Zaveri, J. C. Sackellares, and W. J. Williams, "Modelling of ECoG in temporal lobe epilepsy," *Biomed Sci Instrum*, vol. 24, pp. 187-93, 1988.
[56] K. Lehnertz and C. E. Elger, "Spatio-temporal dynamics of the primary epileptogenic area in temporal lobe epilepsy characterized by neuronal complexity loss," *Electroencephalogr Clin Neurophysiol*, vol. 95, pp. 108-17, 1995.
[57] A. B. Geva, "Feature extraction and state identification in biomedical signals using hierarchical fuzzy clustering," *Med Biol Eng Comput*, vol. 36, pp. 608-14, 1998.
[58] A. Petrosian, R. Homan, S. Pemmaraju, and S. Mitra, "Wavelet based texture analysis of EEG signal for prediction of epileptic seizure," presented at Proceedings of the SPIE - Wavelet Applications in Signal and Image Processing, San Diego, CA, USA, 1995.
[59] A. Petrosian and R. Homan, "The analysis of EEG texture content for seizure prediction," presented at Proceedings of the IEEE Engineering in Medicine and

Biology Society 16th Annual International Conference, Baltimore, MD, USA, 1994.
[60] A. Petrosian, "Kolmogorov complexity of finite sequences and recognition of different preictal EEG patterns," presented at Proceedings of the Eighth IEEE Symposium on Computer-Based Medical Systems, Lubbock, TX, USA, 1995.
[61] B. E. Boser, I. M. Guyon, and V. N. Vapnik, presented at Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory, Pittsburgh, PA, USA, 1992.
[62] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. New York: Cambridge University Press, 2000.
[63] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1999.
[64] K. P. Bennett and C. Campbell, "Support vector machines: hype of hallelujah?," *SIGKDD Explorations*, vol. 2, pp. 1-13, 2000.
[65] C. J. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, pp. 1-47, 1998.
[66] K.-R. M. Muller, S.; Ratsch, G.; Tsuda, K.; Scholkopf, B., "An introduction to kernel-based learning algorithms," *IEEE Transactions on Neural Networks*, vol. 2, pp. 181-201, 2001.
[67] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems*, vol. 13, pp. 18-28, 1998.
[68] E. Osuna, R. Freund, and F. Girosit, "Training support vector machines: an application to face detection," presented at Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico, 1997.
[69] T. Joachims, C. Nedellec, and C. Rouveirol, "Text categorization with support vector machines: learning with many relevant," presented at Machine Learning: ECML-98 10th European Conference on Machine Learning, Chemnitz, Germany, 1998.
[70] P. Hayton, B. Scholkopf, L. Tarassenko, and P. Anuzis, "Support vector novelty detection applied to jet engine vibration spectra," in *Advances in Neural Information Processing Systems 13*, T. K. Leen, T. G. Dietterich, and V. Tresp, Eds. Cambridge, MA: MIT Press, 2001.
[71] V. Wan and W. M. Campbell, "Support vector machines for speaker verification and identification," presented at Proceedings of the 2000 IEEE Signal Processing Society Workshop on Neural Networks for Signal Processing X, Sydney, NSW, Australia, 2000.
[72] S. Mukkamala, G. Janoski, and A. Sung, "Intrusion detection using neural networks and support vector machines," presented at Proceedings of the 2002 International Joint Conference on Neural Networks, Honolulu, HI, USA, 2002.
[73] H. Drucker, D. W., and V. N. Vapnik, "Support vector machines for spam categorization," *IEEE Transactions on Neural Networks*, vol. 10, pp. 1048-1054, 1997.
[74] K.-R. Muller, A. Smola, G. Ratsch, B. Scholkopf, J. Kohlmorgen, and V. N. Vapnik, "Predicting time series with support vector machines," in *Advances in*

Kernel Methods -- Support Vector Learning, B. Scholkopf, C. J. Burges, and A. Smola, Eds. Cambridge, MA: MIT Press, 1999, pp. 211-241.

[75] S. Mukherjee, E. Osuna, and F. Girosi, "Nonlinear prediction of chaotic time series using support vector machines," presented at Proceedings of the 1997 IEEE Workshop on Neural Networks for Signal Processing VII, Amelia Island, FL, USA, 1997.

[76] R. Rosipal and M. Girolami, "An Adaptive Support Vector Regression Filter: A Signal Detection Approach," presented at Proceedings of 9th International Conference on Artificial Neural Networks: ICANN, Edinburgh, UK, 1999.

[77] J. A. K. V. Suykens, J., "Recurrent least squares support vector machines," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 47, pp. 1109-1114, 2000.

[78] M. Pittore, C. Basso, and A. Verri, "Representing and recognizing visual dynamic events with support vector machines," presented at Proceedings of ICIAP '99 - 10th International Conference on Image Analysis and Processing, Venice, Italy, 1999.

[79] A. Garg, I. Cohen, and T. S. Huang, "Adaptive learning algorithm for SVM applied to feature tracking," presented at Proceedings 1999 International Conference on Information Intelligence and Systems, Bethesda, MD, USA, 1999.

[80] N. de Freitas, M. Milo, P. Clarkson, M. Niranjan, and A. Gee, "Sequential support vector machines," presented at Proceedings of the 1999 IEEE Signal Processing Society Workshop on Neural Networks for Signal Processing IX, 1999.

[81] A. Gretton, M. Davy, A. Doucet, and P. J. Rayner, "Nonstationary signal classification using support vector machines," presented at Proceedings of the 11th IEEE Signal Processing Workshop on Statistical Signal Processing, 2001.

[82] M. Davy and S. Godsill, "Detection of abrupt spectral changes using support vector machines: an application to audio signal segmentation," presented at IEEE International Conference on Acoustics, Speech, and Signal Processing, Orlando, FL, USA, 2002.

[83] M.-W. Chang, C.-J. Lin, and R. C. Weng, "Analysis of switching dynamics with competing support vector machines," presented at Proceedings of 2002 International Joint Conference on Neural Networks (IJCNN), Honolulu, HI, USA, 2002.

[84] K. Pawelzik, J. Kohlmorgen, and K.-R. Muller, "Annealed competition of experts for a segmentation and classification of switching dynamics," *Neural Computation*, vol. 8, pp. 340-356, 1996.

[85] G. Ongun, U. Halici, K. Leblebicioglu, V. Atalay, M. Beksac, and S. Beksac, "Feature extraction and classification of blood cells for an automated differential blood count system," presented at Proceedings. IJCNN '01. International Joint Conference on Neural Networks, Washington, DC, USA, 2001.

[86] S. Wang, W. Zhu, and Z.-P. Liang, "Shape deformation: SVM regression and application to medical image segmentation," presented at Proceedings. Eighth IEEE International Conference on Computer Vision, Vancouver, BC, Canada, 2001.

[87] R. A. Reyna, N. Hernandez, D. Esteve, and M. Cattoen, "Segmenting images with support vector machines," presented at Proceedings. 2000 International Conference on Image Processing, Vancouver, BC, Canada, 2000.
[88] O. Chapelle, P. Haffner, and V. N. Vapnik, "Support vector machines for histogram-based image classification," *IEEE Transactions on Neural Networks*, vol. 10, pp. 1055-1064, 1999.
[89] H. Liang and Z. Lin, "Detection of delayed gastric emptying from electrogastrograms with support vector machine," *IEEE Transactions on Biomedical Engineering*, vol. 48, pp. 601-604, 2001.
[90] D. Strauss, G. Steidl, and J. Jung, "Arrhythmia detection using signal-adapted wavelet preprocessing for support vector machines," *Computers in Cardiology*, pp. 497-500, 2001.
[91] J. L. Rojo-Alvarez, A. Arenal-Maiz, and A. Artes-Rodriguez, "Support vector black-box interpretation in ventricular arrhythmia discrimination," *IEEE Engineering in Medicine and Biology Magazine*, vol. 21, pp. 27-35, 2002.
[92] J. Millet-Roig, R. Ventura-Galiano, F. J. Chorro-Gasco, and A. Cebrian, "Support vector machine for arrhythmia discrimination with wavelet transform-based feature selection," *Computers in Cardiology*, pp. 407-410, 2000.
[93] D. Gorur, U. Halici, H. Aydin, G. Ongun, F. Ozgen, and K. Leblebicioglu, "Sleep spindles detection using short time Fourier transform and neural networks," presented at Proceedings of 2002 International Joint Conference on Neural Networks (IJCNN), Honolulu, HI, USA, 2002.
[94] P. Niyogi, C. Burges, and P. Ramesh, "Distinctive feature detection using support vector machines," presented at IEEE International Conference on Acoustics, Speech, and Signal Processing, Phoenix, AZ , USA, 1999.
[95] B. Scholkopf, A. Smola, R. Williamson, and P. Bartlett, "New Support Vector Algorithms," NC2-TR-1998-031, November 1998.
[96] V. N. Vapnik and A. Lerner, "Pattern recognition using generalized portraits," *Avtomatika i Telemekhanika*, vol. 24, pp. 774-780, 1963.
[97] G. Gayraud, "Estimation of functional of density support," *Mathematical Methods of Statistics*, vol. 6, pp. 26-46, 1997.
[98] S. Ben-David and M. Lindenbaum, "Learning distributions by their density levels: A paradigm for learning without a teacher," *Journal of Computer and System Sciences*, vol. 55, pp. 171-182, 1997.
[99] B. Scholkopf, J. Platt, J. Shawe-Taylor, A. Smola, and R. Williamson, "Estimating the support of a high-dimensional distribution," Microsoft Research, Redmond, WA, USA MSR-TR-99-87, 1999.
[100] B. Scholkopf and A. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*. Cambridge, MA, USA: MIT Press, 2001.
[101] B. Scholkopf, "Choice of separating point for nu-SVMs," A. B. Gardner, Ed. Atlanta, GA, USA, 2003.
[102] B. Scholkopf, J. Platt, J. Shawe-Taylor, A. Smola, and R. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol. 13, pp. 1443-1471, 2001.

[103] B. V. Nguyen, "An application of support vector machines to anomaly detection," Ohio University, Athens, OH, USA CS681 Final Report, 2002.
[104] R. Bergman, M. Griss, and C. Staelin, "A personal email assistant," HP Laboratories, Palo Alto, CA, USA HPL-2002-236, 8/30/2002 2002.
[105] L. M. Manevitz and M. Yousef, "One-Class SVMs for Document Classification," *Journal of Machine Learning Research*, vol. 2, pp. 139-154, 2001.
[106] Y. Chen, S. Z. Xiang, and T. S. Huang, "One-class SVM for learning in image retrieval," presented at Proceedings. 2001 International Conference on Image Processing, Thessaloniki, Greece, 2001.
[107] M. Tipping and B. Scholkopf, "A kernel approach for vector quantization with guaranteed distortion bounds," in *Artificial Intelligence and Statistics*, T. Jaakkola and T. Richardson, Eds. San Francisco, CA, USA: Morgan Kaufman, 2001, pp. 129-134.
[108] G. Lanckriet, L. El Ghaoui, and M. I. Jordan, "Robust novelty detection with single-class MPM," presented at Proceedings of the Annual 2002 Neural Information Processing Systems, Vancouver, CA, 2002.
[109] "The Epicentre," vol. 1997: The Neurological Centre, 1997.
[110] G. L. Holmes, *The Comprehensive Evaluation and Treatment of Epilepsy: A practical guide*: Academic Press, 1997.
[111] D. P. White, T. J. Gibb, J. M. Wall, and P. R. Westbrook, "Assessment of accuracy and analysis time of a novel device to monitor sleep and breathing in the home," *Sleep*, vol. 18, pp. 115-26, 1995.
[112] G. G. Haddad, R. A. Epstein, M. A. Epstein, H. L. Leistner, and R. B. Mellins, "The R-R interval and R-R variability in normal infants during sleep," *Pediatr Res*, vol. 14, pp. 809-11, 1980.
[113] G. Brandenberger, J. Ehrhart, F. Piquard, and C. Simon, "Inverse coupling between ultradian oscillations in delta wave activity and heart rate variability during sleep," *Clin Neurophysiol*, vol. 112, pp. 992-6, 2001.
[114] M. H. Bonnet and D. L. Arand, "Heart rate variability: sleep stage, time of night, and arousal influences," *Electroencephalogr Clin Neurophysiol*, vol. 102, pp. 390-6, 1997.
[115] C. James and D. Lowe, pp. 1329-1332, 2000, "Using Independent Component Analysis and Dynamical Embedding to Isolate Seizure Activity in the EEG," presented at Proceedings of the 22nd IEEE Annual International Conference on Engineering in Medicine and Biology Society, 2000.
[116] T.-P. Jung, C. Humphries, T.-W. Lee, S. Makeig, M. J. McKeown, V. Iragui, T. J. Sejnowski, M. Niranjan, E. Wilson, T. Constantinides, and S.-Y. Kung, "Removing electroencephalographic artifacts: comparison between ICA and PCA," presented at Proceedings of the 1998 IEEE Signal Processing Society Workshop on Neural Networks for Signal Processing VIII, Cambridge, UK, 1998.
[117] S. Blanco, H. Garcia, R. Q. Quiroga, L. Romanelli, and O. A. Rosso, "Stationarity of the EEG series," *IEEE Engineering in Medicine and Biology Magazine*, vol. 14, pp. 395-399, 1995.
[118] M. Bozek-Juzmicki, D. Colella, and G. M. Jacyna, "Feature-based epileptic seizure detection and prediction," *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, pp. 564-567, 1994.

[119] J. Gotman, "Automatic detection of seizures and spikes," *J Clin Neurophysiol*, vol. 16, pp. 130-40, 1999.
[120] T. M. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, 1997.
[121] W. B. Pennebaker and J. L. Mitchell, "Probability estimation for the Q-Coder," *IBM J. Res. Develop.*, vol. 32, pp. 737 - 752, 1988.